MONTHLY NOTICES

OF THE

ROYAL ASTRONOMICAL SOCIETY

Vol. 106 No. 5

CONVERSAZIONE ON 1946 OCTOBER 8

A Conversazione was held on the evening of October 8 to commemorate the centenary of the Discovery of Neptune. Professor W. M. Smart gave two addresses on "J. C. Adams and the Discovery of Neptune" during the evening. An exhibition of portraits, manuscripts and books connected with this centenary and also with the tercentenary of the birth of Flamsteed and the quatercentenary of the birth of Tycho Brahe was arranged in the libraries.

MEETING OF 1946 OCTOBER 11

Professor H. H. Plaskett, President, in the Chair

The President referred to the death on the 17th of September of Sir James Jeans and on the 30th of September of Professor H. C. Plummer and read the following resolutions passed by the Council at its meeting that afternoon (the Fellows standing):

"The Council wishes to put on record its sense of the deep loss sustained by the Society, by astronomers generally and by all who are interested in the progress of science, in the death of Sir James Jeans. His contributions to our subject were many. His work on the figures of equilibrium of incompressible and compressible fluids, as developed in his Adams Prize Essay "Problems of Cosmogony and Stellar Dynamics", not only broke new ground in a subject which had engaged the closest attention of Poincaré, Schwarzschild and Darwin, but also formed the starting-point for Jeans's later researches on radiative viscosity, gravitational instability and stellar evolution. His name is associated with that of Rayleigh in the theory of equilibrium radiation, and Jeans himself was responsible for a critical account of the theory of radiation and the quantum theory which appeared as a report to the Physical Society in 1914. He was besides perhaps the greatest of expositors of modern physical science, and will long be remembered for his text books on Electricity and Magnetism and on the Dynamical Theory of Gases, for his more popular books on astronomy and physics, notably the Physical Basis of Music, and for his many lectures.

Sir James Jeans was President of the Society for the years 1925 to 1927, a presidency signalized not only by three brilliant addresses on the award of the gold medal, but also by his endowment of the George Darwin Lectureship."

Also

"The Council desires to record with the deepest regret the loss which the Society has sustained by the death of Henry Crozier Plummer, at one time Royal Astronomer for Ireland and subsequently Professor of Mathematics at the Military College of Science, Woolwich. One of the most modest of men he has left his mark in many branches of our science from dynamical astronomy to the radial distribution of stars in globular clusters, and from the motions of the A-type stars to the optics of astronomical instruments. In addition he was probably the most profound, as he was the most accurate, student of the history of astronomy, and his Halley Lecture in 1942 did much to restore a balanced view of the scientific knowledge and problems of Halley's age.

Professor Plummer was elected to Fellowship in 1899. He served on the Council, and had just commenced his tenure of the presidential chair on the outbreak of the 1939-45 war. That the Society has weathered so well the shock and strains occasioned by that war is due in no small measure to Plummer's wisdom and guidance in these first critical years."

The election by the Council of the following Fellows was duly confirmed:—

William Alexander Calder, M.A., Ph.D., Howard University, Washington, D.C., U.S.A.;

Adolph Hermann Friedmann, LL.D., 15 Frognal Court, London, N.W. 3;

Thomas Alfred Grocock, M.A., B.Sc., The Grammar School, Ripon, Yorkshire;

Carl Keenan Seyfert, M.A., Ph.D., Warner and Swasey Observatory, Cleveland, Ohio, U.S.A.; and

T. V. Viswanathan, M.A., B.Sc., Presidency College, Madras, India.

The following candidates were proposed for election as Fellows of the Society, the names of the proposers from personal knowledge being appended:—

Roland Lebeg Townley Clarkson, The Bungalow, The Avenue, Trimley St. Mary, Ipswich, Suffolk (proposed by W. H. Steavenson);

Arthur William Lane Hall, Devon Lodge, Hartington Road, Sherwood, Nottingham (proposed by W. M. Lindley);

Marjorie Hall Harrison, Yerkes Observatory, Williams Bay, Wis., U.S.A. (proposed by S. Chandrasekhar); and

Guy Ernest Williamson, M.B.E., Hildegarde, Wells Road, Bickley, Kent (proposed by C. S. Wright).

One hundred and thirty-two presents were announced as having been received since the last ordinary meeting, including:—

Strömgren Festschrift (presented by Professor E. Strömgren); and
Atlas der Restlinien, II Band (presented by Rev. Fr. J. Stein, S.J.).

The President welcomed Professor A. Danjon and Dr G. Fayet who were on a visit to attend the centenary celebrations in this country of the discovery of Neptune.

THE SYNTHESIS OF THE ELEMENTS FROM HYDROGEN *

F. Hoyle

(Received 1946 April 6 †)

Summary

Stars that have exhausted their supply of hydrogen in regions where thermonuclear reactions are important enter a collapsing phase. If the mass of the star exceeds Chandrasekhar's limit collapse will continue until rotational instability occurs. Rotational instability enables the star to throw material off to infinity. This process continues until the mass of the remaining stellar nucleus becomes of the order of, or less than Chandrasekhar's limit. The nucleus can then attain a white dwarf equilibrium state.

The temperature generated at the centre of a collapsing star is considered and it is shown that values sufficiently high for statistical equilibrium to exist between the elements must occur. The relative abundances of the elements can then be worked out from the equations of statistical mechanics. These equations are considered in detail and it is shown that a roughly uniform abundance of the elements over the whole of the periodic table can be obtained. The process of rotational instability enables the heavy elements built up in collapsing stars to be distributed in interstellar space.

The results arising from the discussion of the formation of heavy elements lead to a natural explanation of the difference between novae and supernovae.

1. *Introduction*

1.1. *Preliminary remarks.*—The rate at which nuclear re ctions take place in a thermodynamic assembly of material particles and radiation is negligible at ordinary temperatures, except in the case of radioactive nuclei. But as the temperature T increases, the thermal energy of the particles increases until for temperatures in the neighbourhood of 5×10^9 °C. appreciable penetration of the Coulomb barriers of the nuclei by protons and α-particles occurs. Thus, provided the density of the particles is high enough for collisions to be sufficiently frequent the rate of nuclear reactions becomes important at very high temperatures. Indeed for sufficiently large ρ, T the interchanges between nuclei of different atomic weights and nuclear charges become so rapid that statistical equilibrium is set up between the nuclei. The relative abundances of the various elements can then be calculated in terms of ρ, T from the equations of statistical mechanics. Moreover the relative abundances so calculated are independent of the initial composition of the assembly. Thus if we start with an assembly of hydrogen at low temperature and ρ, T are increased until statistical equilibrium occurs the relative abundance will be independent of starting with hydrogen and would have been the same if, for example, we had started with an assembly of helium. The work of Section 5 shows that statistical equilibrium will be set up in a time interval of less than 100 seconds when $\rho = 10^7$ gm. per cm.^{3} and $T = 4 \times 10^9$ °C. This result is very sensitive to the value of T but not to the value of ρ.

* The writer is indebted to a referee for valuable suggestions concerning the preparation of this paper.

† Received in present form 1946 December 2.

No attempt will be made in the present paper to discuss non-statistical conditions of synthesis of the elements from hydrogen. It is, however, of interest to consider whether this restriction will affect the generality of the conclusions reached. In the writer's opinion no such restriction can arise. There are two arguments that support this point of view. Firstly in order to synthesize the heavy elements on the scale required to explain the occurrence of the heavy elements in the Earth it is necessary that a large number of interchanges between atomic nuclei shall take place, and that these interchanges take place over the whole range of atomic weights. This, however, is essentially the condition required for statistical equilibrium. Stated otherwise, the present argument means that if an appreciable fraction of the hydrogen is to be converted into heavy elements then statistical equilibrium must occur (at any rate approximately) since the number of nuclear reactions necessary to bring about the synthesis must be comparable with, or greater than, the number of particles involved (if only a minute fraction of the hydrogen were converted into heavy elements then non-statistical processes such as occur in laboratory nuclear physics could take place). In addition to this general argument there is an important special argument. In order to synthesize the naturally radioactive elements it is necessary to consider reactions that are the reverse of the radioactive disintegrations. Under statistical equilibrium no special difficulty arises, but under non-statistical conditions it seems impossible to proceed (in a reverse direction) along the radioactive series beyond such extremely short-lived nuclei as *Ra C′* and *Th C′*.

The above remarks suggest the problem of finding values of ρ, T such that the statistical equations give the relative abundances of the elements observed in nature. The phrase "in nature" requires some discussion. The information available concerning the distribution of the elements may be summarized as follows:—

(1) It is known that a large quantity of material occurs both in the Galaxy and in the extragalactic nebulae in the form of a diffuse interstellar gas. Although the amount of this material is not yet accurately determined, there is important observational evidence in favour of the view that the total mass of interstellar material in a nebula exceeds the combined masses of the stars by an appreciable factor. The interstellar material is believed to consist very largely of hydrogen, the quantity of heavy elements amounting to only a few per cent. at most. The heavy elements are present partly in gaseous form and partly as interstellar dust.

(2) As a result of a fairly thorough analysis of the structure of the stars it is now known that a large fraction of the combined masses of the stars is due to hydrogen. In certain stars hydrogen contributes at least 70 per cent., whereas at the other end of the scale stars are known to occur (see Section 2) in which hydrogen is almost entirely absent. This diversity in hydrogen content is due to the interplay of two processes with opposing tendencies:

(i) The generation of energy within a star due to thermonuclear reactions has the net effect of synthesizing helium from hydrogen (in stars of mass less than about half the solar mass $M_\odot$ the helium is commonly believed to be synthesized directly from the hydrogen, whereas in stars of mass greater than about $M_\odot/2$ the synthesis occurs through the carbon-nitrogen cycle). Thus the effect of nuclear reactions *within a normal star* is to increase the helium concentration

at the expense of the hydrogen (the distinction between a normal star and a collapsing star is discussed in detail in Section 2). The synthesis from hydrogen of elements heavier than helium is negligible in normal stars * in which the central temperatures never appreciably exceed 5×10^7 °C.

(ii) The loss of hydrogen by synthesis of helium is offset in some stars by rapid accretion of interstellar hydrogen.†

Accordingly we form the following picture of the development of a star during the normal phase of its history.‡ The star condenses from the diffuse interstellar material § and has initially the same composition as the interstellar material. As a result of thermonuclear reactions hydrogen is converted into helium. In some stars the loss of hydrogen is offset by a rapid accretion of further quantities of interstellar material.

(3) Additional data concerning the composition of the stars is obtained from the study of stellar spectra. Although the identification of elements over the whole periodic table is not yet completed, the results so far obtained provide strong confirmation of the view that all elements known on the Earth are also present in the stars. It has been estimated by Russell || that in the surface layers of the stars hydrogen is at least a thousand times as abundant as all the metals combined.

(4) By far the most detailed evidence concerning the distribution of heavy elements comes from the direct study of the contents of the earth's crust.¶ The distribution so obtained differs markedly from the astrophysical evidence discussed above in that only a small quantity of hydrogen and an almost negligible amount of helium is present on the Earth. The most common elements in the crust are silicon, magnesium, and oxygen.

In the present paper we are not so much concerned with the abundance of isolated elements as with suitably averaged values. We may group the elements into a number of ranges of atomic weight, e. g. the ranges 1 to 14 inclusive, 15 to 24 inclusive etc., and average the abundance of all the elements falling within each group. When this is done an averaged abundance is obtained for each range of atomic weight. These averaged abundances rarely fall below one part in a million (uranium forms about one part in 10^5 and lead about five parts in 10^6, so that this statement holds up to the heaviest elements in the periodic table). Now although a difference in abundance between two elements by a factor of 10^3 may be of very great economic importance (one of the elements being regarded as "common" and the other as "rare") such a difference is unimportant from the standpoint of the present paper since factors arise in the present discussion that may be as large as 10^{200}. From the work given below it will appear that a remarkable feature of the survey of elements present in the Earth's crust is the very small variation in abundance that occurs over the whole range of atomic weights.

The above remarks concerning the uniformity of distribution of the elements in the Earth's crust are strengthened by noticing that the heaviest elements would

* H. A. Bethe, *Phys. Rev.*, **55**, 434, 1939.
† F. Hoyle and R. A. Lyttleton, *M.N.*, **101**, 227, 1941.
‡ See Section 2.
§ F. Hoyle, *M.N.*, **105**, 302, 1945.
|| H. N. Russell, *Ap. J.*, **78**, 239, 1933.
¶ V. M. Goldschmidt, *Skr. norske Vidensk Akad.*, 1937.

sink towards the centre when the Earth was in a liquid state. Thus the proportion of elements with atomic weights greater than 40 in the Earth as a whole may well be considerably higher than is indicated by a study of the crust alone. Results obtained from observations of the propagation of sound waves through the body of the Earth confirm this statement. It is found* that the density of material near the centre of the Earth is close to twelve. Allowing for increase of density arising from the high pressure in the inner regions of the Earth it has been estimated that such material would have a density of about eight under normal laboratory conditions. This means that the central regions of the Earth are occupied by material with a density close to that of iron or copper. It is usually inferred from this result that the core of the Earth is occupied by elements with atomic weights lying between 50 and 70. The core may however contain ten per cent. by mass of the heaviest elements without any disagreement with observation arising.

It will be seen from the above remarks that the evidence available concerning the distribution of the elements in the universe is very fragmentary. Nevertheless the astrophysical data are sufficient to bring out the important result that hydrogen is by far the most abundant element. This conclusion gives strong support to the view that it should be possible to trace back the history of the universe to a state in which hydrogen is the only element present. That is, it must be possible to show how the various elements have been synthesized from hydrogen. The present paper attempts to show how such a synthesis takes place. Stated more explicitly we suppose:

(i) Initially the only element present in the universe is hydrogen.

(ii) Helium is synthesized by thermonuclear reactions taking place in "normal" stars.

(iii) A further process occurs that synthesizes higher elements from hydrogen and helium. The elements produced are regarded as having a distribution similar to that found on the Earth.

The object of the present paper is to discuss (iii).

1.2. *Relation with previous investigations.*—The equations of statistical equilibrium between atomic nuclei have been discussed by a number of authors.† The most recent work is by Chandrasekhar and Henrich (referred to below as the C-H theory). The C-H theory deals with the relative abundance of the elements at temperatures between 8×10^9 and 6×10^9 °C. and at densities of order 10^7 gm. per cm.^{3}. Under these conditions only about one part in a million by mass is converted into elements of atomic weight greater than that of helium. The relative abundances of the elements are well reproduced in the range between *C* and *Cl*, but beyond atomic weight 40 the abundances given by the C-H theory become quite negligible. Thus the C-H theory fails to account for the existence of elements beyond argon. One of the main features of the present work is that the required abundances will be obtained over the whole of the periodic table. The chief points of difference between the present and previous discussion may be summarized as follows:—

(1) Account is taken of the degeneracy of the electrons at high densities.

* H. Jeffreys, The Earth, p. 220, Cambridge, 1929; *M.N., Geophys. Suppl.* **4**, 62, 1937.

† T. E. Sterne, *M.N.*, **93**, 736, 1933; R. H. Fowler, Statistical Mechanics, p. 655, Cambridge, 1936; S. Chandrasekhar and L. R. Henrich, *Ap. J.*, **95**, 288, 1942.

(2) A large fraction of the material is converted into elements heavier than helium (as opposed to one part in a million in the C-H theory).

(3) The inclusion of electron degeneracy enables the required relative abundances to be obtained over the whole range of the periodic table without the temperature of the assembly being appreciably greater than 10^{10} °C.

In addition to these technical points the present work differs from the C-H theory on a number of questions of principle:

(*a*) It is not sufficient to find the values of ρ, T required for an assembly to contain the proper relative abundances of the elements. It is also necessary that ρ, T can be converted to low density and temperature (such as occurs on the Earth or in interstellar space) *without the relative abundances being changed.* This means that the mixture must remain frozen during some cooling-expansion process, for if statistical equilibrium were maintained throughout the process the relative abundances would undergo important changes.

(*b*) It is necessary to find the *place in the universe* where the required values of ρ, T occur.

In addition to previous investigations dealing with the synthesis of elements there are also a number of papers concerned with the properties of matter at high temperatures and densities that the writer has read with interest.*

1.3. *Procedure to be followed.*—The chief aims of the present paper may be summarized briefly:—

(1) To find values of ρ, T such that the equations of statistical equilibrium give the required abundances of the elements over the whole range of the periodic table.

(2) To show that a cooling-expansion process can be found such that the values of ρ, T satisfying (1) can be reduced to low density and temperature without the composition of the mixture being changed.

(3) To find a place in the universe where the values of ρ, T satisfying (1) must occur and to explain in a satisfactory way the origin of the cooling-expansion process.

There are two ways of discussing this programme. The indirect way is to construct the equations of statistical equilibrium and to search by trial and error until values of ρ, T are obtained that satisfy (1). The next step would be to use these values of ρ, T in a search for processes satisfying (2), (3) and (4). The second and more direct way is to discuss (3) and to find what values of ρ, T satisfy (2) also. Then having arrived at a pair of values of ρ, T (or possibly at more than one pair of values) we can use (1) and (2) as consistency checks on the theory. The latter procedure will be adopted in the following work. It is shown below that consideration of (2) and (3) leads to two possible ranges of values of ρ, T. One of these ranges yields the abundance of the elements of atomic weight up to about 80 (this is material described as case (1) in Section 8), whilst the other case yields the required abundances over the remainder of the periodic table (this material is described in Section 8 as case (2)).

An important by-product of the present investigation is a natural explanation of the difference between novae and supernovae. The theory of supernovae here given is similar in a number of respects to the neutrino-emission theory

* W. Baade and F. Zwicky, *Phys. Rev.*, **45**, 138, 1934 and **46**, 76, 1934 ; L. Landau, *Nature*, **141**, 333, 1938 ; F. Zwicky, *Phys. Rev.*, **55**, 726, 1939 ; F. Cernushi, *Phys. Rev.*, **56**, 450, 1939.

given by Gamow and Schoenberg.* The chief point of difference is that whereas in the Gamow-Schoenberg theory removal of energy occurs only through neutrino emission, in the present theory a more rapid rate of removal of energy arises from nuclear transformations.

2. *Some General Properties of Stars*

2.1. *Normal Stars.*—In the present section we shall consider stars in which nuclear transformations do *not* abstract thermal energy from stellar material (as will be seen in later sections important cases can occur where the absorption of energy by nuclear reactions becomes the dominant process in determining the evolution of a star). Under these conditions the only important way in which a star loses energy is by radiation from the surface. It follows from this, as will be proved at the end of the present subsection, that the star must satisfy the hydrostatic equation

$$\Re \frac{d}{dr}\left(\frac{\rho T}{\mu\beta}\right) = -\frac{GM(r)\rho}{r^2} \tag{1}$$

to a very high degree of approximation, where $\Re$ is the gas constant; μ, ρ, T are the mean molecular weight, density and temperature of the stellar material; β is the ratio of the gas pressure $\Re\rho T/\mu$ to the total pressure $\Re\rho T/\mu + aT^4/3$, which includes the pressure $aT^4/3$ due to radiation, where a is Stefan's constant; G is the gravitational constant, r is a radial coordinate that measures distance from the centre of the star, and $M(r)$ is the mass of the star contained in a sphere of radius r.

Now when a star satisfies (1) either accurately or approximately the central temperature T_c cannot differ by a factor of much more than two from the value †

$$\frac{G\mu\beta_c M}{3\Re R}, \tag{2}$$

where M is the total mass of the star, β_c is the value of β at the centre, and R is the radius of the star (more accurately R should be taken such that $M(R) \simeq M/2$, because it is possible for a star to possess an outer envelope of large radius but containing only a small fraction of the mass if μ is appreciably less in the envelope than in the interior ‡). The temperature of the surface of the star is small compared with (2) and there must be an outward temperature gradient of order

$$\frac{G\mu\beta_c M}{3\Re R^2}$$

in the star. This temperature gradient will result in there being an outward flux of energy from inside the star to the surface. The magnitude of this energy flux follows immediately from the well known equation of radiative equilibrium and is of order

$$\frac{16\pi caR}{3\kappa\bar{\rho}^2(1-X^2)}\left(\frac{G\mu\beta_c M}{3\Re R}\right)^{7.5}, \tag{3}$$

where κ is the coefficient of opacity, c is the velocity of light, X is the fraction

* G. Gamow and M. Schoenberg, *Phys. Rev.*, **59**, 539, 1941 ; G. Gamow, *Phys. Rev.*, **59**, 617, 1941 and **65**, 20, 1944.

† A. S. Eddington, I.C.S., p. 92, Cambridge, 1930.

‡ F. Hoyle and R. A. Lyttleton, *M.N.*, **102**, 218, 1942.

by mass of hydrogen in the star, and $\bar{\rho}$ the mean density of material in the star is given by

$$\frac{4}{3}\pi R^3\bar{\rho} = M. \tag{4}$$

Substituting for $\bar{\rho}$ in (3) leads to the following expression for the energy flux:

$$\frac{3ca}{\pi\kappa(1-X^2)} \cdot \left(\frac{G\mu\beta_c}{3\mathfrak{R}}\right)^{7\cdot5} \cdot M^{5\cdot5}R^{-0\cdot5}. \tag{5}$$

The expression (5) not only gives the outward flux of energy in the star but also gives the rate of radiation from the surface. Thus, if the rate of radiation from the surface is greater than (5), the surface temperature must fall, since the loss of energy at the surface cannot be compensated by the flux of energy from inside the star. Similarly, if the radiation from the surface is less than (5), the surface temperature must rise, since the supply of energy from inside the star exceeds the loss by radiation.

If thermonuclear reactions taking place in the star generate sufficient energy to compensate exactly for the loss of energy at the surface, then the star remains in equilibrium. If the generation of energy exceeds the rate of loss then the star will expand. As will be seen from (2) the temperature in the energy-producing regions falls with the expansion, and this has the effect of rapidly reducing the rate of production of energy (the carbon nitrogen cycle gives an energy production approximately proportional to ρ^2T^{18}, whilst direct synthesis helium from hydrogen gives an energy generation approximately proportional to $\rho^2T^{4\cdot5}$). Thus the star expands until the energy production falls to a value equal to the loss at the surface. If, on the other hand, the generation of energy due to the synthesis of helium from hydrogen is less than the loss at the surface, then the star will contract and the temperature will rise in accordance with (2). Thus in this case, provided the quantity of hydrogen present in the energy-producing regions is not negligible, the star will contract until energy production becomes equal to the loss at the surface (the discussion given in 2.2 shows that the present proviso is very important). We may conclude that stars containing an appreciable amount of hydrogen will adjust their dimensions (and hence their central temperatures) in such a way that the energy generation is just sufficient to balance the loss at the surface, which is given approximately by (5).

Now the generation of energy by the carbon nitrogen cycle for example (similar remarks apply to direct synthesis of helium from hydrogen) can be written in the approximate form *

$$4\pi E\int_0^R \rho^2T^\eta r^2\,dr, \tag{6}$$

provided the central temperature is not too far from the equilibrium value, where η varies nearly as $M^{-1/12}$ and is about 18 for the Sun, and E is an energy production constant. The integral (6) can be written as

$$\frac{4}{3}\pi E\lambda R^3\bar{\rho}^2T_c^\eta,$$

where λ is an averaging constant that has effectively the same value from one star

* F. Hoyle and R. A. Lyttleton, *M.N.*, **102**, 177, 1942.

to another. Substituting for $\bar{\rho}$ from (4) and for T_c from (2) gives an energy production

$$\frac{3E\lambda}{4\pi}\left(\frac{G\mu\beta_c}{3\mathfrak{R}}\right)^{\eta}\frac{M^{\eta+2}}{R^{\eta+3}}. \tag{7}$$

For equilibrium conditions the expressions (5) and (7) are equal and R is given approximately by

$$\left\{\frac{E\lambda\kappa(1-X^2)}{4ac}\right\}^{\frac{2}{2\eta+5}}\cdot\left(\frac{G\mu\beta_c}{3\mathfrak{R}}\right)^{\frac{2\eta-15}{2\eta+5}}\cdot M^{\frac{2\eta-7}{2\eta+5}}. \tag{8}$$

Thus, since it can be shown that β_c can be expressed in terms of μ and M, it follows that R can be calculated * explicitly in terms of μ and M. The value of R so determined may be substituted in (5) to give the rate of emission L at the surface in terms of μ and M. Thus *both* R and L can be calculated explicitly in terms of μ and M. These remarks refer to stars in equilibrium. Such stars are *normal* in the sense used in the Introduction.

In a refined discussion of the structure of stars the approximate expressions (5) and (7) have to be replaced by accurate values (the expressions (5) and (7) are approximate in the numerical constants multiplying them and not in the dependence on M, R and μ).† For the present purposes the above discussion is preferable because a number of points are brought out that tend to be obscured in a more detailed theory. Moreover the present investigation shows what results must of necessity turn up in a thorough-going analysis. For example, it is seen from (8) that $R \propto E^{2/(2\eta+5)}$ and from (5) that $L \propto E^{-1/(2\eta+5)}$, so that both R and L are very insensitive to the value of the energy production constant. Furthermore we see that this result is inherent in the problem and cannot be removed by more detailed calculations.

Apart from a numerical factor the expression (5) is identical with a well-known formula due to Eddington.

At this stage we may dispose of the asumption made at the beginning of the present subsection. That is, we have to show that a collapsing star which *only loses energy through radiation at the surface* must satisfy (1) to a high degree of approximation. The total thermal energy contributed by the material of a star is of order $\mathfrak{R}MT_c$. Now the star cannot contract appreciably until an amount of energy has been radiated at the surface that is of the same order as $\mathfrak{R}MT_c$. Thus the time required for appreciable contraction to occur is of order $\mathfrak{R}MT_c/L$. To obtain an idea of the magnitude of this quantity put $M = M_\odot = 2 \times 10^{33}$ gm., $T_c = 2 \times 10^7$ °C., $L = L_\odot = 4 \times 10^{33}$ ergs per sec. Then the time required for appreciable contraction to occur is of order $2{\cdot}8 \times 10^7$ years. This time may be compared with the time that would be required for collapse if (1) were not satisfied to a good approximation. The latter time is easily seen to be of order $\sqrt{(R^3/GM)}$. For $R = R_\odot = 7 \times 10^{10}$ cm. and $M = M_\odot$, $\sqrt{(R^3/GM)}$ is about $1{\cdot}6 \times 10^3$ sec. Although for different M, T_c, R the numerical values will differ from those given here it is clear that in all cases

$$\frac{\mathfrak{R}MT_c}{L} \gg \sqrt{\left(\frac{R^3}{GM}\right)}.$$

---

* F. Hoyle and R. A. Lyttleton, *loc. cit.*

† F. Hoyle and R. A. Lyttleton, *loc. cit.*

This proves the required result that (1) must be satisfied to a good approximation when the only loss of energy is by radiation at the surface.

2.2. *Collapsing Stars.*—In this subsection we consider a collapsing star in which the rate of generation of energy is less than the rate of radiation from the surface. Two different cases may be distinguished:

(1) The star contains an appreciable fraction by mass of hydrogen but the internal temperature is too low to give sufficient energy generation for equilibrium. This case occurs during the condensation of a star from interstellar material, and the collapse continues until the temperature rises to a sufficiently high value for equilibrium to occur.

(2) The generation of energy by synthesis of helium from hydrogen may cease, due to the exhaustion of the supply of hydrogen in the energy-producing regions (hydrogen may still be present in the atmosphere). This will occur in a period of about 10^8 years (compared with an age of the universe of about 10^{10} years) for stars of large mass and luminosity unless the hydrogen is replenished by accretion of interstellar hydrogen. Thus, if accretion ceases to be important * for such a star, the supply of hydrogen will become exhausted in a period of time that is short compared with the age of the universe. The star will then enter a period of contraction. As the radius decreases the central temperature will increase in accordance with (2). The rising temperature will remove the last vestiges of hydrogen present in the inner parts of the star and the star if sufficiently massive must collapse until rotational instability occurs.† The condition on the mass is that it must appreciably‡ exceed Chandrasekhar's limit $5{\cdot}75M_\odot/\mu_e^2$. Sufficient material is thrown off to infinity during the instability process for the mass of the remaining stellar nucleus to be reduced to a value of the order of, or less than, Chandrasekhar's limit.

From the point of view of the present paper two different cases may be distinguished. These two cases may be examined by discussing the instability process in further detail.

(i) The temperature rises sufficiently high for nuclear reactions to become important *before* the onset of rotational instability (a temperature of about 4×10^9 °C. is required).

(ii) A state of rotational instability is reached before nuclear reactions become important.

Owing to the conservation of angular momentum, any initial rotation possessed by a star increases as the star contracts. During contraction the rotational velocity is proportional to the reciprocal of the radius of the star. Thus the centrifugal force acting on the stellar material increases as the inverse cube of the radius. It follows that, since the gravitational force increases only as the inverse square of the radius, the ratio of centrifugal force to gravitational force must increase inversely as the radius. When the centrifugal force becomes comparable with the gravitational force, instability occurs. If we regard equality

* F. Hoyle, *M.N.*, **105**, 363, 1945.

† F. Hoyle, *M.N.*, in course of publication.

‡ The quantity μ_e is defined somewhat differently from μ. Thus whereas μ is given by the product of the mass of the hydrogen atom and the number of particles per gram of material, the quality μ_e is defined as the product of the mass of the hydrogen atom and the number of free electrons per gram of material. For the stars considered in the present paper μ_e is always close to 2.

of these two forces as an approximate criterion for rotational instability, then we can estimate the degree of contraction required to produce instability for various values of the initial equatorial rotational velocity. For example, consider a star of mass $10M_{\odot}$ that exhausts its supply of hydrogen and enters the period of contraction discussed above. Before contraction begins, the values of ρ, T at the centre of the star would be about 3 gm. per cm.^{3} and $3{\cdot}7 \times 10^7$ °C., whilst the radius would be about $3{\cdot}5 \times 10^{11}$ cm. To a sufficient approximation we may suppose that the star contracts through a series of homologous configurations, so that the central values of ρ, T are proportional to the inverse cube and the inverse first power respectively of the radius. The following table gives the various quantities that are of interest for a set of values of the initial rotational velocity :—

TABLE I

The Effect of the Initial Rotational Velocity on Collapsing Stars

	Mass $=10M_{\odot}$. Initial radius $=3{\cdot}5 \times 10^{11}$ cm.						
Initial rotational velocity (km. per sec.).	1	5	10	15	20	40	100
Factor by which radius must be reduced to make centrifugal and gravitational forces equal.	$3{\cdot}82 \times 10^5$	$1{\cdot}53 \times 10^4$	$3{\cdot}82 \times 10^3$	$1{\cdot}70 \times 10^3$	$9{\cdot}52 \times 10^2$	$2{\cdot}38 \times 10^2$	$3{\cdot}82 \times 10$
Initial central temperature in °C.	$3{\cdot}5 \times 10^7$	$3{\cdot}5 \times 10^7$	$3{\cdot}5 \times 10^7$	$3{\cdot}5 \times 10^7$	$3{\cdot}5 \times 10^7$	$3{\cdot}5 \times 10^7$	$3{\cdot}5 \times 10^7$
Central temperature at time of rotational instability in °C.	$>4 \times 10^9$	$>4 \times 10^9$	$>4 \times 10^9$	$>4 \times 10^9$	$>4 \times 10^9$	$>4 \times 10^9$	$1{\cdot}34 \times 10^9$
Initial central density in gm. per cm.^{3}.	3	3	3	3	3	3	3
Central density at onset of rotational instability in gm. per cm.^{3}.	$1{\cdot}67 \times 10^{17}$	$1{\cdot}07 \times 10^{13}$	$1{\cdot}67 \times 10^{11}$	$1{\cdot}47 \times 10^{10}$	$2{\cdot}59 \times 10^9$	$4{\cdot}07 \times 10^7$	$1{\cdot}67 \times 10^5$

The values given in Table I are necessarily approximate (in particular relativity effects are neglected in the first set of values), but even so they show clearly that unless the initial rotational velocity of the star is of the order of, or greater than, 100 km. per sec. the temperature will rise to 4×10^9 °C. before rotational instability occurs. Now at a temperature of 4×10^9 °C. nuclear reactions become important and we have case (i). Indeed it follows from Table I that, if the initial rotational velocity is of the order of, or less than, 40 km. per sec., we have case (i), whereas, if the initial rotational velocity is of the order of, or greater than, 100 km. per sec., we have case (ii). Thus, since the initial rotational velocity will vary from one star to another, it follows that some stars will satisfy case (i) and others case (ii). In the present paper we are only concerned with stars in which the initial rotational velocity is small enough for case (i) to apply. The range of values of the initial rotational velocity used in the following work is roughly 10 to 40 km. per sec.

The large densities appearing in the above table suggest that electron degeneracy may arise during the contraction of the star, and that the contribution

of the pressure of the degenerate electrons may become important. As will be seen in a later section, this is the case, but only *after* nuclear reactions have become important. Before nuclear reactions become important, the density increases at the same rate as the cube of the temperature, so that the ratio of gas pressure to radiation pressure remains constant during this phase of the contraction. Now, as Chandrasekhar has shown *, pressure degeneracy does not become important unless β_c is greater than a value close to 0·908. From the known relation † between β_c and M, it can be shown that β_c will not exceed 0·9 unless $M\mu_e^2/M_\odot < 5{\cdot}75$. In the following discussion we shall only consider stars satisfying $M\mu_e^2/M_\odot > 5{\cdot}75$, so that degeneracy does not arise before nuclear reactions become important. If a star undergoes appreciable contraction after the temperature has risen to 4×10^9 °C., then degeneracy will always occur. This is due to nuclear reactions which prevent the temperature from rising appreciably above 10^{10} °C. At this stage β_c increases, because the gas pressure continues to increase on account of the increasing density but radiation pressure remains effectively constant. If sufficient contraction occurs, β_c will rise above 0·908 and the stellar material will become degenerate.

2.3.—The remarks of 2.2 provide a suitable basis for a theory of the synthesis of the elements. It has been shown that the temperatures arising in collapsing stars that have exhausted their supply of hydrogen become sufficiently high for statistical equilibrium between atomic nuclei to occur provided

(1) The mass of the star appreciably exceeds Chandrasekhar's limit.

(2) The initial rotational velocity before collapse begins is less than about 40 km. per sec.

Thus our first object of finding a place in the universe where the elements may be synthesized has been achieved. Furthermore, a cooling-expansion process of the type discussed in the Introduction is also available in the process of rotational instability. This serves not only to reduce the density and temperature but to distribute the elements in interstellar space.

In view of the importance of the processes discussed in 2.2 it is desirable to seek independent confirmation of these results. It is therefore fortunate that observational evidence can be cited in support of the starting-point of the present investigations. Thus what appears to be an excellent example of the break up of a rotationally unstable star has recently been investigated in detail by Baade and Minkowski.‡ In this particular case it is possible to observe directly the nebulous material that was thrown off by the supernova of 1054 A.D. The mass of this material is found to be of order 15 $M_\odot$ so that the star must have thrown off an appreciable proportion of its original mass. The estimate given by Minkowski for the radius of the remaining stellar nucleus confirms that the material in the nucleus is very dense, and thereby provides support for the view that the outburst occurred in a collapsing star whose supply of hydrogen had become exhausted. Thus, if we take the mass in the nucleus as $M_\odot$, which is probably an underestimate, together with Minkowski's estimate of about 10^9 cm. for the radius, the present mean density is about 5×10^5 gm. per cm.^{3}. Allowing for a central condensation of about 20 (the central density in

* S. Chandrasekhar, Introduction to the Study of Stellar Structure, p. 434, Chicago, 1938.

† F. Hoyle and R. A. Lyttleton, *loc. cit.*

‡ W. Baade, *Ap. J.*, **96**, 188, 1942 ; R. Minkowski, *Ap. J.*, **96**, 199, 1942.

a star is about 20 times the mean density), this gives a present central density in the nucleus of about 10^7 gm. per cm.^{3}. We may note further that the central density before outburst would be unlikely to be less than 10^7 gm. per cm.^{3}. Thus, taking 16 $M_\odot$ as a rough estimate of the initial mass, it follows that the radius at outburst could not be much greater than $2{\cdot}5 \times 10^9$ cm. If this value of the radius is used in (2), together with $\mu\beta_c = 1$, an estimate of $3{\cdot}4 \times 10^9$ °C. is obtained for the central temperature before outburst. This value is just of the order that is of interest in the present problem. Accordingly it is seen that there is not only theoretical evidence but also observational data in favour of the model used in the following sections of the present paper.

3. *The Equations of Statistical Equilibrium*

The next step is to construct the equations of statistical equilibrium between atomic nuclei (these equations applying at temperatures exceeding 4×10^9 °C.). Write n_P, n_N, n_A^Z for the number of free protons per cm.^{3}, the number of free neutrons per cm.^{3}, and the number of atomic nuclei with atomic weight A and charge Z per cm.^{3}, respectively. Then under conditions of statistical equilibrium n_A^Z can be expressed in terms of the three quantities n_P, n_N and T. These equations are well known and will be quoted here.* In Fowler's notation the required equations are

$$\left.\begin{aligned} Vn_P &= \xi \frac{\partial}{\partial \xi}\left\{\sum_q \ln\left(1+\xi e^{-\epsilon_q/kT}\right)\right\}, \\ Vn_N &= \zeta \frac{\partial}{\partial \zeta}\left\{\sum_r \ln\left(1+\zeta e^{-\epsilon_r/kT}\right)\right\}, \\ Vn_A^Z &= \chi \frac{\partial}{\partial \chi}\left\{\sum_s \pm \ln\left(1\pm\chi e^{-\epsilon_s/kT}\right)\right\}, \end{aligned}\right\} \tag{9}$$

where V is the volume occupied by the assembly and ϵ_q, ϵ_r, ϵ_s represent the energy states (energy of translation together with internal excitation) of the proton, neutron and nucleus of atomic weight A and charge Z, respectively. The $\pm$ sign depends on A; if A is odd the nuclei obey Fermi-Dirac statistics and the plus sign must be taken, whilst if A is even the nuclei obey Einstein-Bose statistics and the minus sign must be used. The parameters ξ, ζ, χ appear in using the theorem of steepest descents and they satisfy the relations

$$\left.\begin{aligned} \chi &= \xi^Z \zeta^{(A-Z)} \exp\left(-Q_A^Z/kT\right), \\ Q_A^Z &= c^2\{m_A^Z - Zm_P - (A-Z)m_N\}, \end{aligned}\right\} \tag{10}$$

m_P, m_N, m_A^Z being the masses of the proton, neutron and nucleus of weight A and charge Z, respectively. There is a set of equations of the form (9) and (10) for each pair of values of A, Z.

The above equations can be simplified in two ways. First, to the degree of accuracy aimed at in the present paper the contribution to ϵ_s of the excited states of the nuclei can be neglected (the errors arising from this step are not greater than a factor of about 10^2 in the relative abundances of the nuclei. This is small compared with the factors that arise from other considerations). Second,

* R. H. Fowler, *op. cit.*, p. 655. The symbol " ln " denotes thel ogarithm to the base e, and the symbol " log " will be used to denote logarithms to the base 10.

even at the highest densities discussed below ($\simeq 10^{11}$ gm. per cm.^{3}), the parameters ξ, ζ, χ are sufficiently small compared with unity for it to be possible to use "classical" statistics for the protons, neutrons and heavy nuclei. These simplifications enable the sums in (9) to be written as *

$$\left.\begin{aligned} \sum_q \ln(1+\xi e^{-\varepsilon_q/kT}) &= \xi V(2\pi m_P kT)^{\frac{3}{2}}/h^3, \\ \sum_r \ln(1+\zeta e^{-\varepsilon_r/kT}) &= \zeta V(2\pi m_N kT)^{\frac{3}{2}}/h^3, \\ \pm\sum_s \ln(1\pm\chi e^{-\varepsilon_s/kT}) &= \chi V(2\pi m_A^Z kT)^{\frac{3}{2}}/h^3. \end{aligned}\right\} \quad (11)$$

Equations (11) together with (10) give

$$\ln n_A^Z = Z\ln n_P + (A-Z)\ln n_N + (A-1)\ln\{h^3/(2\pi kT)^{\frac{3}{2}}\} - \frac{Q_A^Z}{kT} - \frac{3}{2}Z\ln m_P - \frac{3(A-Z)}{2}\ln m_N + \frac{3}{2}\ln m_A^Z. \quad (12)$$

This equation has been expressed in a convenient numerical form by Chandrasekhar and Henrich †, who use 10^9 °C. as the unit of T, and the millimass unit ($10^{-3}\ m_P$) as the unit of Q_A^Z. With these units, and putting $m_P = m_N$, equation (12) becomes

$$\log n_A^Z = 34{\cdot}08 + \frac{3}{2}\log T + \frac{3}{2}\log A + \frac{4{\cdot}73}{T}\cdot Q_A^Z + A\left(\log n_N - 34{\cdot}08 - \frac{3}{2}\log T\right) - Z\log\left(\frac{n_N}{n_P}\right). \quad (13)$$

Equation (13) gives one of the basic equations of the present problem. The remaining equations are rather less straightforward in their derivation and will be treated in greater detail. A relation can be derived between n_N/n_P and n_e, where n_e is the electron density. This relation is due to interchange between an electron and a proton on the one hand and a neutron on the other hand. The interchange arises from β-processes taking place among the nuclei. The neutrinos and antineutrinos (according to the usual formulation of the properties of these particles) emitted in the β-processes will escape from the star and will remove energy in the manner discussed by Gamow and Schoenberg.‡ Provided the interchanges are sufficiently frequent (we shall assume this to be the case), statistical equilibrium will be set up, which will be similar in character to the case of a simple dissociating gas. In the latter problem we have atoms of type A_1 combining with atoms of type A_2 to form molecules A_1A_2, together with the reverse reaction in which A_1A_2 dissociates into the separate atoms, A_1 and A_2. This can be represented symbolically as

$$A_1 + A_2 \rightleftarrows A_1A_2.$$

The solution of this problem is known § and can be taken over to the proton, neutron, electron case by regarding the neutron as the "molecule" and the proton and electron as the "atoms". The equations determining the equilibrium are

* R. H. Fowler, *op. cit.*, p. 655.
† Chandrasekhar and Henrich, *op. cit.*
‡ Gamow and Schoenberg, *loc. cit.*
§ Fowler, *op. cit.*, p. 157.

the first two equations of (9) together with

$$\left.\begin{aligned} Vn_e &= \lambda \frac{\partial}{\partial \lambda}\left\{\sum_t \ln\left(1+\lambda e^{-\epsilon_t/kt}\right)\right\}, \\ \zeta &= \lambda\xi \exp\left(-\mathcal{J}/kT\right), \\ \mathcal{J} &= c^2(m_N - m_P - m_e), \end{aligned}\right\} \tag{14}$$

where ϵ_t represents the energy levels of the electron and m_e is the electron mass. Putting ζ, ξ equal to the values given by (9) and (11) gives

$$\frac{n_N}{n_P} = \lambda\left(\frac{m_N}{m_P}\right)^{\frac{3}{2}} \exp\left(-\frac{\mathcal{J}}{kT}\right) \simeq \lambda e^{-\mathcal{J}/kT}. \tag{15}$$

The first of equations (14) determines λ in terms of n_e. This equation is comparatively easy to solve for both the cases $\lambda \ll 1$ and $\lambda \gg 1$, but is rather awkward when $\lambda \simeq 1$. Throughout the following work the form of solution in which $\lambda \gg 1$ will be used. This is necessary at the highest densities considered (of order 10^{11} gm. per cm.^{3}) but is only a moderate approximation at the lowest densities (of order 10^7 gm. per cm.^{3}) treated in the work of later sections. This procedure is however fully justified by the circumstances that in calculations of the relative abundances of the elements the terms depending on λ are unimportant at the lowest densities (10^7 gm. per cm.^{3}) but are very important at the highest densities (10^{11} gm. per cm.^{3}). Thus our approximation is very good in the range of density where the terms in λ matter, whilst the approximation is only moderate for the range of density in which the terms in λ are unimportant. When $\lambda \gg 1$ and account is taken of the relativistic variation of mass of the electrons in forming the energy levels ϵ_t, the first of equations (14) can be solved to give*

$$\left.\begin{aligned} \ln\lambda &= m_e c^2 \sqrt{(1+x^2)}/kT, \\ x &= (n_e/5{\cdot}87 \,.\, 10^{29})^{\frac{1}{3}}, \end{aligned}\right\} \tag{16}$$

the temperature being here measured in °C. Equations (15) and (16) give

$$\left.\begin{aligned} \ln\left(\frac{n_N}{n_P}\right) &= \frac{m_e c^2}{kT}\left\{\sqrt{(1+x^2)} - \frac{\mathcal{J}}{m_e c^2}\right\} = \frac{m_e c^2 y}{kT}, \\ \text{where} \quad y &= \sqrt{(1+x^2)} - \mathcal{J}/m_e c^2 = \sqrt{(1+x)^2} - 0{\cdot}51. \end{aligned}\right\} \tag{17}$$

If we again measure T in units of 10^9 °C., the first of equations (17) becomes

$$\log \frac{n_N}{n_P} = \frac{4{\cdot}73}{T} \,.\, 0{\cdot}543y. \tag{18}$$

Substituting in (13) gives

$$\log n_A^Z = 34{\cdot}08 + \frac{3}{2}\log T + \frac{3}{2}\log A + \frac{4{\cdot}73}{T}(Q_A^Z - 0{\cdot}543yZ)$$
$$+ A(\log n_N - 34{\cdot}08 - \tfrac{3}{2}\log T). \tag{19}$$

The second of equations (16), equations (18) and (19) give the complete scheme of statistical equations in the present problem.

A more useful form of the equation determining the parameter x can be

* S. Chandrasekhar, *op. cit.*, p. 393 ; R. H. Fowler, *op. cit.*, p. 652.

obtained by expressing n_e in terms of the density of the material. If we write ρ' for the density excluding the contribution of the free neutrons (in the following discussion the free protons never make an appreciable contribution to the density, but at the highest densities considered below it turns out that the free neutrons do make an appreciable contribution), then to a sufficient approximation we can write

$$n_e \simeq \frac{\rho'}{2m_P} \tag{20}$$

for material not containing hydrogen. This equation follows from the consideration that the assembly must be effectively electrically neutral. This requires that n_e shall be equal to the total positive charge density of the atomic nuclei, which is close to $\rho'/2m_P$ together with a contribution equal to the density of the positrons present in the assembly. The latter term can be calculated from statistical mechanics * and it turns out that its value can be neglected in comparison with $\rho'/2m_P$. Except when ρ' is higher than 10^{10} gm. per cm.^{3}, it is sufficient to put ρ' equal to the total density ρ. The equation for x becomes

$$\rho' = 1{\cdot}98 \times 10^6 x^3. \tag{21}$$

Equation (21) enables x and therefore y to be determined in terms of ρ'. Thus, since Q_A^Z is a known quantity, it follows from (19) that n_A^Z is determined in terms of ρ', T and n_N.

4. *The Helium Zone*

4.1. It would now be possible to return to the physical model discussed in Section 2 and to apply the statistical equations of Section 3. Before doing this, however, it is more convenient to discuss the further properties of these equations and to return to the physical model at a later stage. The two questions that will be considered in the present section and in Section 5 respectively are:

(1) If we regard T, ρ' as Cartesian coordinates, then in the T, ρ'-plane a curve can be drawn with the following important property:—Material at values of T, ρ' represented by any point on one side of the curve is composed almost entirely of helium, whereas material at any point, T, ρ' on the other side of the curve is composed almost entirely of heavy elements, the main mass of the elements in the latter case having atomic weight greater than 50.

(2) Throughout Section 2 it was assumed that nuclear reactions take place sufficiently rapidly when $T > 4 \times 10^9$ °C., $\rho' > 10^7$ gm. per cm.^{3} for the statistical equations to be applicable, whereas when $T < 4 \times 10^9$ °C. it was supposed that the mixture remains frozen. This question will be discussed in detail in Section 5.

It will now be shown when statistical equilibrium occurs that, for a given density ρ', there is a closely determinable value of T in the neighbourhood of which the material changes over from being almost entirely composed of helium to being almost entirely composed of heavy elements. It has to be shown further that in the latter case the most abundant element has atomic weight greater than 50. It is most convenient to assume the latter result to begin with, and then to prove it at the end of the discussion.

Equation (19) can be used to express n_4^2 in terms of y (i.e. ρ'), T, and n_N.

* R. H. Fowler, *op. cit.*, p. 653, equation (1830).

Thus we have

$$\log n_4^2 = 34{\cdot}08 + \frac{3}{2}\log T + \frac{3}{2}\log 4 + \frac{4{\cdot}73}{T}(Q_4^2 - 1{\cdot}086y)$$

$$+ 4(\log n_N - 34{\cdot}08 - \frac{3}{2}\log T). \qquad (22)$$

Equations (19) and (22) enable n_N to be eliminated and give the following relation between n_A^Z and n_4^2:—

$$\frac{\log(An_A^Z) - 34{\cdot}08 - 3/2\log T - 5/2\log A}{A} = \frac{\log 4n_4^2 - 34{\cdot}08 - 3/2\log T - 5/2\log 4}{4}$$

$$+ \frac{4{\cdot}73}{T}\left\{\frac{Q_A^Z}{A} - \frac{Q_4^2}{4} + 0{\cdot}543\,y\left(\frac{1}{2} - \frac{Z}{A}\right)\right\}. \qquad (23)$$

Consider now the orders of magnitude of the three terms in equation (23).

(i) *Second term on right-hand side of equation* (23). The quantity Q_A^Z/A is the packing fraction of atoms of atomic weight A and charge Z. The values of the various quantities involved are shown in Table II for a number of elements, using experimental values for the packing fraction.*

TABLE II

	He^4	O^{16}	Si^{28}	Fe^{56}	Cu^{63}	Kr^{82}	Sn^{118}	Pb^{208}
Q_A^Z/A	7·57	8·54	9·01	9·27	9·27	9·22	9·12	8·35
Z/A	0·5	0·5	0·5	0·46	0·46	0·44	0·42	0·40
$Q_A^Z/A - Q_4^2/4$	0·0	0·97	1·44	1·70	1·70	1·65	1·55	0·78
$Q_A^Z/A - Q_4^2/4 + 0{\cdot}543y(1/2 - Z/A)$	0·0	0·97	1·44	1·70 $+0{\cdot}02y$	1·70 $+0{\cdot}02y$	1·65 $+0{\cdot}034y$	...	...

These values taken together with the following values of y for various ρ' show that the second term on the right-hand side of (23) is of order unity when $A > 16$ and $4 < T < 10$:—

TABLE III

ρ' (gm. per cm.^{3})	10^7	10^8	10^9	10^{10}	10^{11}
x	1·72	3·70	7·96	17·2	37·0
y	1·48	3·32	7·51	16·6	36·5

It will be remembered that the form of the term in y arises from the formulae for relativistically degenerate electrons. The values given in Tables II and III confirm the statement made in Section 3 that the term in y only becomes important when the use of these formulae gives a good approximation for the electron distribution (the condition for the approximation to be a good one is that $x \gg 1$).

(ii) *First term on the right-hand side of* (23).—If all the material were in the form of helium, then

$$\log 4n_4^2 = \log\left(\frac{\rho'}{m_P}\right) = \log\rho' + 23{\cdot}78,$$

whilst if other elements are present

$$\log 4n_4^2 < \log\rho' + 23{\cdot}78.$$

* O. Hahn, S. Flugge and J. Mattauch, *Phys. Z.*, 1941. The value used here for Kr^{82} is less than the value given by these authors. The present value has been chosen to fit smoothly into a maximum of the packing fraction between 60 and 70.

It follows therefore that this term has a magnitude also of order unity when $\rho' < 10^{11}$ gm. per cm.^{3} and $T > 4$.

(iii) *Left-hand side of* (23).—We now use the assumption that the most abundant heavy element occurs for $A > 50$. Consider for example a most abundant heavy element of atomic weight 60. Table IV gives the values of $\log(60\, n_{60}^Z)$ for a number of specified values of the left-hand side of (23).

TABLE IV

Left-hand side of (23)	$\log(60\, n_{60}^Z)$
0·0	$34{\cdot}08 + 3/2 \log T + 5/2 \log 60$
−0·1	$28{\cdot}08 + 3/2 \log T + 5/2 \log 60$
−0·2	$22{\cdot}08 + 3/2 \log T + 5/2 \log 60$

It follows from these numerical values that, when the left-hand side of (23) is zero, $60\, n_{60}^Z > \rho'/m_P$ for all $\rho' < 10^{11}$ gm. per cm.^{3} but that, when this term is $-0{\cdot}2$ then $60\, n_{60}^Z$ is extremely small compared with ρ'/m_P for all $\rho' > 10^6$ gm. per cm.^{3}. Accordingly only a very small change in this term is required to change the mixture from being composed almost entirely of heavy elements to being composed almost entirely of helium. This result, taken in conjunction with the fact that only a very small change in T is required to change the second term on the right-hand side of (23) by an amount of order 0·1, shows that the change over from helium to heavy elements must occur at values of T close to that given by putting the right-hand side of (23) equal to zero. Moreover for the purpose of calculating this value of T it is sufficiently accurate to put

$$\log 4n_4^2 = \log \rho' + 23{\cdot}78.$$

(This equation holds to a good approximation when the mixture is almost all helium, and it can also be used as an order of magnitude relation during the changeover from helium to heavy elements, but it could not be used when T has become sufficiently small for the mixture to be composed almost entirely of heavy elements.) This gives

$$T = \frac{18{\cdot}92\{Q_A^Z/A - Q_4^2/4 + 0{\cdot}543y(0{\cdot}5 - Z/A)\}}{10{\cdot}3 + 3/2 \log T + 5/2 \log 4 - \log \rho'}. \qquad (24)$$

When T is greater than the value given by (24), the mixture is composed almost entirely of helium and, when T is less than (24), the mixture is composed almost entirely of heavy elements.

A further question arises as to what values of A and Z should be used in (24). It is clear that the values of A, Z giving the largest $(Q_A^Z - 0{\cdot}543yZ)/A$ will give the highest T. This means that the region of heavy elements in the T, ρ'-plane is most extensive when the maximum value of $(Q_A^Z - 0{\cdot}543yZ)/A$ with respect to *both* A and Z is used in (24). It follows from Table II that this will occur for $A > 50$, which confirms the statement made above that the most abundant heavy element occurs for $A > 50$. The value of $(Q_A^Z - 0.543yZ)/A$ for Cu^{63} gives the following table of values of T, ρ' satisfying (24):—

TABLE V

ρ' (gm. per cm.^{3})	10^6	10^7	10^8	10^9	10^{10}	10^{11}
T (in units of 10^9 °C.)	4·7	5·5	6·6	8·6	11·2	17·4

These values are sufficiently accurate for the present purpose but some approximation is involved on account of the way $\log T$ is treated in making

these calculations. The required values are insensitive to the value of log T and much arithmetical labour is saved by putting log T equal to a representative constant (1·0 in the present case). A second order correction is easily inserted after the first order values have been worked out. The pairs of values of T, ρ' given in Table V lie on the curve in Fig. 1 which separates the helium zone from the zone of heavy elements.

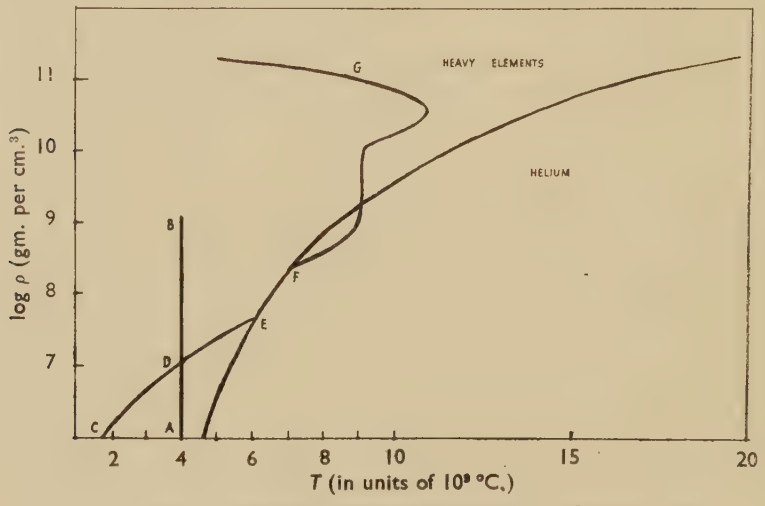


FIG. 1.—*The T, ρ' diagram. The track DEFG is for the case $\alpha = 0{\cdot}24$.*

4.2. *The values of n_N, n_P in the helium zone.*—In the helium zone $\log(4n_4^2) = \log\rho' + 23{\cdot}78$ to good approximation. The value of n_N can be determined in terms of T, ρ' from

$$\log n_N = 34{\cdot}08 + \frac{3}{2}\log T + (\log\rho' - 10{\cdot}3 - \frac{3}{2}\log T - \frac{5}{2}\log 4) - \frac{4{\cdot}73}{T}(7{\cdot}57 - 0{\cdot}271y), \qquad (25)$$

which is derived by putting $Q_4^2/4 = 7{\cdot}57$ and $\log 4n_4^2 = \log\rho' + 23{\cdot}78$ in (22). When n_N has been calculated from (25), the value so obtained can be used in (18) to find n_P. In this way the following numerical values of n_N, n_P were obtained:—

TABLE VI

$\rho' = 10^7$ gm. per cm.³. $\log n_4^2 = 30{\cdot}18$

T	6	7	8	9	10
$\log n_N$	28·1	29·0	29·7	30·2	30·6
$\log n_P$	27·5	28·4	29·2	29·8	30·2

$\rho' = 10^8$ gm. per cm.³. $\log n_4^2 = 31{\cdot}18$

T	8	9	10
$\log n_N$	30·2	30·7	31·1
$\log n_P$	29·1	29·7	30·2

$\rho' = 10^9$ gm. per cm.³. $\log n_4^2 = 32{\cdot}18$

T	9	10
$\log n_N$	31·5	31·9
$\log n_P$	29·4	30·0

The remarkable feature of Table VI is the large equilibrium values of n_N, n_P that occur for T close to 10. This result is of importance in the discussions of Sections 6 and 9. It will be noticed that at $T=10$ the mass of the neutrons is comparable with, but less than, the mass of the helium. This is the reason why the density of the neutrons was not included in ρ' (the quantity ρ' appears in the formulae only through the relation $n_e \doteq \rho'/2m_P$ so that ρ' must evidently not include the neutrons).

4.3. *The values of n_N, n_P in the heavy element zone.*—Equation (19) can be written in the form

$$\log n_N = 34{\cdot}08 + \frac{3}{2}\log T - \frac{4{\cdot}73}{T}\left(\frac{Q_Z^A}{A} - \frac{0{\cdot}543yZ}{A}\right) + \frac{1}{A}\left\{\log (An_A^Z) - 34{\cdot}08 - \frac{3}{2}\log T - \frac{5}{2}\log A\right\}. \qquad (19')$$

If this equation is applied to the element of maximum abundance then we have

(i) For $10^6 < \rho' < 10^{11}$ the last term on the right-hand side of (19′) may be neglected, since we can put $\log(An_A^Z) = \log\rho' + 23{\cdot}78$.

(ii) The heavy element of maximum abundance occurs when $(Q_A^Z - 0{\cdot}543yZ)/A$ is a maximum with respect to *both* A and Z. The full discussion of this question is left over until Section 7. We may anticipate the result that, when $\rho' \leqslant 10^9$ gm. per cm.^{3}, the required maximum value can be obtained simply by examining the experimentally determined values of Q_A^Z (such as are given in Table II). It then appears that the maximum occurs for A near 60. Thus provided $\rho' \leqslant 10^9$ gm. per cm.^{3} reliable estimates for n_N can be obtained by inserting the experimental values of Q_A^Z and Z/A for Cu^{63} in (19′) (copper serves as an example of an element with atomic weight near 60). The values of n_N so determined can be used in (18) to give n_P. The following values were calculated in this way:—

TABLE VII

$\rho' = 10^7$ gm. per cm.^{3}

T	3	4	5
$\log n_N$	20·8	24·5	26·7
$\log n_P$	19·5	23·5	26·0

$\rho' = 10^8$ gm. per cm.^{3}

T	3	4	5	6
$\log n_N$	21·5	25·0	27·2	28·6
$\log n_P$	18·7	22·9	25·5	27·2

$\rho' = 10^9$ gm. per cm.^{3}

T	3	4	5	6	7	8
$\log n_N$	23·1	26·2	28·1	29·4	30·3	31·1
$\log n_P$	16·7	21·4	24·3	26·2	27·6	28·6

5. *The Conditions for Statistical Equilibrium*

In this section the conditions under which the equations of Section 3 can be applied will be briefly examined. The necessary requirements in the present problem are:

(i) Energy must be statistically distributed among the energy states of translation of each type of particle present (this condition is necessary for the evaluation of the sums in equation (9)).

(ii) Many important nuclear reactions involve the emission of γ-rays (the reactions in which helium is synthesized from hydrogen for example). Statistical equilibrium requires that there shall be detailed balancing between these reactions and the reverse reactions that involve the absorption of a γ-ray. This will only be the case if there is thermodynamic equilibrium between matter and radiation. Furthermore the energies of the γ-rays involved in the nuclear reactions will in some cases be large compared with kT (for the highest temperatures used in the present paper kT is only about 1 M.V.). Thus we require equilibrium to extend up to quanta with energies large compared with kT.

(iii) There must be suitable chain reactions connecting any two pairs of values of A, Z, provided that the nuclei with these pairs of values occur in appreciable abundance. As an example of such a chain reaction, the following known reactions connecting O^{16} and Al^{27} may be noted:

$$O^{16}+He^4 \rightleftarrows F^{19}+H^1, \qquad F^{19}+H^1 \rightleftarrows Ne^{20}+h\nu,$$
$$Ne^{20}+He^4 \rightleftarrows Na^{23}+H^1, \qquad Na^{23}+H^1 \rightleftarrows Mg^{24}+h\nu,$$
$$Mg^{24}+He^4 \rightleftarrows Al^{27}+H^1.$$

Unfortunately there is not sufficient laboratory evidence available to give a known chain reaction for every pair of values of A, Z. Two difficulties arise in attempting to apply only known reactions to the present problem. First, there is far too little published evidence concerning nuclei with atomic weights greater than 50, and, second, at the high densities involved in the present problem it is to be expected that triple collisions between nuclear particles will be important, whereas in the laboratory densities are far too low for these processes to be noticeable. In these circumstances it has been decided to make the following assumptions concerning the chain reactions :—

(*a*) That a chain reaction is available between any two pairs of values of A, Z and that only neutron, proton, and α-particles reactions are required to establish the chain (these reactions include $n-\alpha$, $n-p$, $n-\gamma$, $n-2n$, $p-\alpha$, $p-\gamma$, $p-n$, $\alpha-n$, reactions together with their inverses).

(*b*) That for 5 M.V. protons* and α-particles the cross-sections of all reactions occurring in these chains exceed 10^{-28}.

(*c*) That statistical equilibrium cannot be established without the proton reactions, which are required not only to enable the chain reactions to take place but also to satisfy condition (iv) given below.

(iv) The equilibrium between protons and neutrons requires a sufficiently rapid β-process to occur. The importance of a particular β-reaction depends not only on the lifetime for the process but also on the abundance of the nucleus involved.

The conditions (i), (ii), (iii) will now be considered in further detail, while the discussion of condition (iv) will be postponed until the results of Sections 6, 7 and 8 are available. The first question to decide is the length of time that can be allowed for statistical equilibrium to be established. The work of Section 9 may be anticipated which shows that condition (iv) taken in conjunction with the properties of collapsing stars fixes a time-scale of about 100 sec. for the most rapid stages in the collapse of a star (it will be seen that the early stages of collapse are much slower than this).

Condition (i).—This requirement will be satisfied provided all types of particle experience an appreciable number of collisions per second. It is easy to show

* 1 M.V.≡million electron volts$=1{\cdot}59\times10^{-6}$ergs.

that this must be the case. Taking firstly the charged particles, the collision cross-section is of order 10^{-16} and the number of collisions experienced by a particle is of order $10^{-16}\,vn$ per sec., where v, n are suitably averaged values of the velocity and particle density respectively. For the values of T, ρ' under consideration, $v \simeq 10^9$ cm. per sec., $n > 10^{29}$, so that the number is of order 10^{22} per sec. Thus condition (i) is satisfied by a very large margin for these particles.

The collision cross-section for neutrons is much smaller than for charged particles. A rough estimate of 10^{-24} for the elastic collision cross-section of the neutrons still gives a collision rate, however, of 10^{14} per sec. Thus even in this case the number of collisions is amply large enough for statistical equilibrium to be established among the states of kinetic energy of the neutrons.

Condition (ii).—To establish this condition it has to be shown that a sufficiently large interchange of energy takes place between matter and radiation. Moreover this interchange must take place up to energies large compared with kT. The phrase "sufficiently large" may be taken to have the following precise meaning:—that the amount of thermal energy of the material converted into radiant energy per sec. per cm.^{3} in any frequency range ν to $\nu + d\nu$ must be large compared with the equilibrium energy density of radiation given by the Planck formula

$$\frac{8\pi h\nu^3 d\nu}{c^3\{\exp(h\nu/kT)-1\}}. \tag{26}$$

Under these conditions detailed balancing will occur and equal quantities of radiant energy will be converted into thermal energy. It is to be expected that it will be most difficult to establish this requirement when $h\nu \gg kT$, and this case will now be considered.

The number of electrons with energy greater than E_0 per cm.^{3} is given by

$$\frac{2n_e}{\pi^{\frac{1}{2}}(kT)^{\frac{3}{2}}}\int_{E_0}^{\infty}\exp(-E/kT)E^{\frac{1}{2}}\,dE,$$

where n_e is the number of electrons per cm.^{3}. (This statistical equilibrium formula can be used since condition (i) is satisfied.) For the case $E_0 \gg kT$, this expression can be written to sufficient accuracy as

$$2n_e\left(\frac{E_0}{\pi kT}\right)^{\frac{1}{2}}\exp\left(-\frac{E_0}{kT}\right). \tag{27}$$

Now there is an appreciable probability that an electron will lose almost the whole of its kinetic energy in the Bremstrahlung emission of a single quantum.* For collisions with a nucleus of charge Z the cross-section for this process is about $5\times 10^{-27}Z^2$. Combining this with (27), it is easy to see that the number of quanta with energy greater than E_0 emitted by the electrons per cm.^{3} per sec. is of order

$$10^{-26}n_e v_e\left(\sum_{A,Z}Z^2n_A^Z\right)\left(\frac{E_0}{kT}\right)^{\frac{1}{2}}\exp\left(-\frac{E_0}{kT}\right), \tag{28}$$

where v_e is the velocity of electrons with energy E_0, and the summation is taken over all nuclei present in the material. Now the number of quanta with energy greater than E_0 per cm.^{3} under thermodynamic equilibrium is

$$\frac{8\pi}{c^3}\int_{\nu_0=E_0/h}^{\infty}\frac{\nu^2 d\nu}{\exp(h\nu/kT)-1} \simeq \frac{8\pi E_0^2 kT}{c^3h^3}\exp\left(-\frac{E_0}{kT}\right), \tag{29}$$

* W. Heitler, *Quantum Theory of Radiation*, p. 170, Oxford, 1936.

when $E_0 \gg kT$. An order of magnitude estimate for the time required for equilibrium to be set up can now be obtained by dividing (29) by (28). If we put $A=4$, $Z=2$, $2n_e=n_4^2=10^{30}$ per cm.^{3}; $v_e=10^{10}$ cm. per sec., $E_0=20$ M.V., $kT=0{\cdot}3$ M.V., this gives a time of order 10^{-12} sec. Thus condition (ii) is also satisfied by a very large margin.

Condition (iii).—Proton, α-particle and neutron reactions will now be considered subject to the assumptions (*a*), (*b*), and (*c*) made above. Take proton reactions first. Since condition (i) is satisfied the number of protons per cm.^{3} with energy lying between E and $E+dE$ is

$$\frac{2n_P E^{\frac{1}{2}} dE}{\pi^{\frac{1}{2}}(kT)^{\frac{3}{2}}} \cdot \exp\left(-\frac{E}{kT}\right).$$

Thus the probability of a given nucleus of atomic weight A and charge Z experiencing a proton reaction in a time δt is

$$\frac{2\delta t n_P}{\pi^{\frac{1}{2}}(kT)^{\frac{3}{2}}} \int_0^\infty \sigma_A^Z(E) v_P(E) \exp(-E/kT) E^{\frac{1}{2}} dE, \tag{30}$$

where $\sigma_A^Z(E)$ is the proton reaction cross-section of the given nucleus, and $v_P(E)$ is the velocity of protons of energy E. The reaction cross-section increases rapidly with E (for the range of E that makes an appreciable contribution to the integral (30)) on account of the Coulomb barrier of the nucleus.

Experimental cross-sections are nearly always measured at one or two discrete values of E, and it is not possible to use measured values over the whole range of E in (30). This integral can be estimated so far as order of magnitude is concerned, however, by taking $E^{\frac{1}{2}} . v_P(E) . \sigma_A^Z(E)$ outside the integral at a particular value E_0 of E, where E_0 is chosen so that $\sigma_A^Z(E)$ is known from experiment. (This remark assumes that the reaction under consideration is one for which experimental measurements are available. As pointed out in the following paragraph, only a small proportion of the reactions of interest in the present problem fall in this class.) To avoid overestimating the integral a lower limit of E_0 is then used instead of zero. That is, (30) can be estimated as

$$2\delta t n_P . v_P(E_0) \sigma_A^Z(E_0) . \left(\frac{E_0}{\pi kT}\right)^{\frac{1}{2}} \exp\left(-\frac{E_0}{kT}\right). \tag{31}$$

In the majority of reactions of importance in the present problem no experimentally determined cross-sections are available and it is necessary to fall back on the assumption (*b*) made above. According to this assumption when $E_0=5$ M.V. the cross-section $\sigma_A^Z(E) \geqslant 10^{-28}$ for all A, Z. An estimate for the probability of proton reactions can then be obtained by inserting these values in (31). (If we put $\sigma_A^Z(E_0)=10^{-28}$ the probability will be underestimated. It is desirable to underestimate the probabilities rather than to overestimate them.) It follows immediately from (31) that a nucleus A, Z will have a proton reaction in a time of less than 100 seconds provided n_P is greater than the values given in the following table:—

TABLE VIII

T	4	5	6	7	8
$\log n_P$	22·2	21·0	20·2	19·5	19·2
$\log n_4^2$	22·5	21·3	20·5	19·8	19·5

The assumption (*b*) also gives a cross-section $\geqslant 10^{-28}$ when $E_0=5$ M.V. for

α-particle reactions, so that by allowing for the difference between the thermal velocities of the protons and α-particles we obtain immediately the corresponding criterion for n_4^2, in order that a nucleus A, Z shall experience an α-particle reaction in a time of less than 100 seconds.

The cross-sections for neutron reactions are of the order of, or less than, 10^{-24}. Moreover the neutrons do not have to penetrate a Coulomb barrier, so that the incident neutrons may have kinetic energies of order kT. Thus the probability of neutron reactions is far higher than the probability of proton reactions (this follows since n_N is always greater than n_P).

It was assumed throughout Section 2 that statistical equilibrium applies when $T > 4 \times 10^9$ °C. but not when $T < 4 \times 10^9$ °C. This statement follows immediately by comparing the values of $\log n_P$ in Table VII with the requirements for statistical equilibrium given in Table VIII. It is seen that the value of n_P becomes too small as T decreases below 4×10^9 °C. for condition (iii) to hold. This result is insensitive to the value of ρ' and may be applied over the range $10^6 < \rho' < 10^8$ gm. per cm.^{3}.

Condition (iv).—This condition is discussed below in Section 8.

6. *The Track of the Stellar Material in the Helium Zone*

6.1. In Sections 3, 4 and 5 the equations of statistical equilibrium have been set up and certain of their general properties discussed. It is now proposed to return to the model of a collapsing star and to apply the results of Sections 3, 4 and 5 during the collapsing phase. It will be assumed throughout the present and following sections that a state of rotational instability has not been reached. In this connection it will be recalled that the onset of rotational instability can be arranged to occur at any stage we please by suitably adjusting the initial rotational velocity (see Table I).

At any particular time the values of T, ρ' for the material at the centre of a collapsing star can be represented as a point in the T, ρ'-plane. This point moves with time and the work of the present and following sections enables its track to be found.

The equations of statistical equilibrium can be applied when $T > 4 \times 10^9$ °C. (During the early stages of contraction the time available for equilibrium to be set up is much greater than the value of 100 seconds mentioned in Section 5. Thus statistical conditions may occur before T rises to 4×10^9 °C. *during the contraction*. On the other hand during expansion after rotational instability has taken place a time of several seconds only is available for establishing equilibrium. Thus on expansion the mixture will remain frozen when T falls below 4×10^9 °C. This question is discussed in detail in Section 8.) For the star of mass 10 $M_\odot$ considered in Table I the value of ρ' is about 5×10^6 gm. per cm.^{3} when $T = 4 \times 10^9$ °C., while for stars of other masses the corresponding values of ρ' are given approximately by $5 \times 10^7 (M_\odot/M)$ gm. per cm.^{3} (the central density of a "normal" star is approximately proportional to $1/M$). Thus the values of ρ', for the mass $M < 50\ M_\odot$ but exceeding Chandrasekhar's limit, lie in the range $10^6 < \rho' < 5 \times 10^7$ gm. per cm.^{3} at the time T reaches 4×10^9 °C. It follows therefore that the material at the centre of a collapsing star crosses the line $T = 4$ (T now in units of 10^9 °C.) at a point lying between A and B in Fig. 1. For the sake of definiteness a star will be considered in the following work that crosses this line

at $\rho' = 10^7$ gm. per cm.^{3} (marked D in Fig. 1). This requires a star of mass close to 5 $M_\odot$. To the left of the line $T = 4$ it follows from (2) (which applies in this region of the T, ρ'-diagram) that the track of the material at the centre of the star is given by

$$\rho' = 10^7 (T/4)^3 \text{ gm. per cm.}^3. \tag{32}$$

This portion of the track of the material is shown in Fig. 1 by the curve CD.

From Fig. 1 it is seen that on passing the point D (moving from left to right in the figure) the stellar material enters the heavy element zone. Before reaching D the material is composed largely of helium, but on passing D the helium is converted almost entirely into heavy elements (it will be recalled that the helium arises by synthesis from hydrogen; this process occurring while the star is in a "normal" equilibrium state). Now the synthesis of heavy elements from helium gives an important energy yield, since the packing fractions for the heavy elements are larger than the packing fraction of helium. Indeed it is easy to see from the values of the packing fraction given in Table II that the total amount of energy released in the conversion of the helium must be of order $Mc^2/500$. This value may be compared with the total energy generated by the synthesis of helium from hydrogen during the "normal" equilibrium state. This latter energy is of order $Mc^2/125$, so that the conversion of helium to heavy elements near the point D in Fig. 1 yields about 1/4 of the total energy generated while the star was in a "normal" state. This amount is large enough for a new "collapsed" equilibrium state to be set up, in which the material at the centre of the star is in the neighbourhood of the point D in the T, ρ'-diagram. The lifetime of the star in this new equilibrium state can only be a small fraction, however, of the lifetime in the normal state. This is due partly to the smaller total energy yield and partly to the increased rate of radiation from the surface (this is shown by (5) which is proportional to $R^{-1/2}$ so that the rate of radiation is considerably larger in the collapsed equilibrium state than in the normal state). In addition there is an important third reason explained in the following paragraph that reduces the lifetime of the collapsed equilibrium state by an appreciable factor.

When the lifetime of the collapsed equilibrium state is over, the material at the centre of the star will pass the point D in the T, ρ'-diagram, and further collapse of the star (further collapse must occur when energy generation ceases) will lead to the material evolving further along the track given by (32). Now at the point E in Fig. 1 the material reaches the helium zone, which means that there must be a *reconversion* of the material from heavy elements back into helium. In order to effect this reconversion energy must be supplied to the material in an amount equal to the energy that was obtained on passing the point D. There are two processes that contribute to this energy:

(1) The collapse of the star yields gravitational energy.

(2) Since the central regions of a star have a temperature higher than material at greater distances from the centre it follows that the first material in the star to pass the point D in the T, ρ'-diagram will be material lying in the central regions. This remark taken with the fact that once the point D has been passed the evolution from D to E will be comparatively rapid, shows that material at the centre will have reached E before an appreciable fraction of the remaining material of the star has passed D. Thus owing to the star not being at a uniform temperature the process of conversion of helium to heavy elements and the reverse process

of reconversion back to helium will be mixed together, and a proportion of the energy released in the first process will be available for use in the reverse process.

Although it is certain that gravitational energy is required for the reconversion of the heavy elements back to helium it is not possible to estimate the amount quantitatively without calculating the amount of energy that is available through the process (2) of the previous paragraph. This is difficult and will not be attempted. For the present purpose it is sufficient to know that gravitational energy is required before the stellar material can enter the helium zone. Thus when the track of the material reaches E in the T, ρ'-diagram, gravitational energy must be supplied before the track can enter the helium zone. Now the value of ρ' increases whilst the gravitational energy is being supplied, so that the point in the T, ρ'-diagram (representing the material) must move during the time that this energy is made available. It follows therefore that during the supply of the gravitational energy the track of the material *must move along the curve that divides the heavy element zone from the helium zone.* This portion of the track is marked EF in Fig. 1. The important result follows immediately that since the curve EF does not satisfy (32) the hydrostatic equation of support (1) cannot now be satisfied even as an approximation. Stated in more physical terms, when the point E is reached the temperature cannot increase at a sufficiently rapid rate for (1) to be approximately satisfied because the gravitational energy supplied by contraction is swallowed up by the reconversion of heavy elements to helium. We may anticipate the results of Section 9 by stating that the evolution of the material along EF only takes a time of several seconds. Thus the collapse of the star becomes extremely swift when the point E is reached. As will be shown in 6.2 and Section 7 this rapid collapse is also maintained beyond F. Thus the point E is of the nature of a critical point; before E is reached the collapse of the star is very slow (occupying a time of several thousand years at least) whilst beyond E the collapse is catastrophic, the whole evolution then being carried out in a time of a few seconds.

6.2. The next step in the investigation is to trace the track of the material through the helium zone. In this work we shall be concerned with densities that rise appreciably above 10^9 gm. per cm. These high densities necessitate an important change in the formula for the internal energy of the material. Up to now it has been sufficiently accurate to take the thermal energy of the material as $3\mathfrak{R}\rho T/2\mu$ per cm.^{3}. This expression together with the energy density aT^4 of the radiation gave the total energy per cm.^{3} used in the above work. At sufficiently high densities, however, the quantity $3\mathfrak{R}\rho T/2\mu + aT^4$ becomes appreciably less than the energy of degeneracy of the electrons. The latter energy is given by the expression *

$$2\pi m_e^4 c^5 x^4/h^3, \tag{33}$$

when $x \gg 1$ (powers of x smaller than x^4 being neglected). When (21), which expresses x in terms of ρ', is used (33) takes the form

$$1{\cdot}53 \times 10^{15}(\rho')^{\frac{4}{3}}. \tag{33'}$$

It is seen that (33') exceeds $3\mathfrak{R}\rho' T/2\mu$ when $\rho'/\mu = 10^9$ gm. per cm.^{3} $T \leqslant 10^{10}$ °C. The contribution (33') to the energy of the material arises from the Pauli exclusion principle which requires the average energy of the degenerate electrons to be large

* S. Chandrasekhar, *op. cit.*, p. 392.

when the density is high (this is quite apart from the value of the temperature T, and would apply even if T were zero). In the following discussion the ordinary thermal energy of the material and the energy of radiation will both be neglected in comparison with (33′).

The results given in Table VI show that T can only rise to 10^{10} °C. if n_N takes a value comparable with n_4^2. Thus in order for T to rise to 10 in the T, ρ'-diagram an appreciable quantity of helium must be converted into free neutrons. Now a large quantity of energy has to be supplied to the stellar material to make this conversion possible. This energy can only be derived from the gravitational energy supplied by the collapse of the star. Thus the situation is similar to that occurring between the points E and F. That is, the energy supplied by contraction is largely absorbed in nuclear reactions and only a small proportion goes into raising the temperature of the material. Between E and F the energy is absorbed by the reconversion of heavy elements to helium, whereas inside the helium zone the energy is used up in producing free neutrons. This is the reason why the rapid collapse of the star between F and E must continue after the material has entered the helium zone.

An order of magnitude estimate for the amount of energy released in the collapse of the star can be obtained in terms of ρ' and M. The "energy released" in the present sense is meant to include energy that is available for absorption in any of the following three ways * :—

(1) In increasing the value of T. This energy not only includes the increased thermal energy of the material and the increased energy of radiation but also the loss of energy by neutrino emission in the β-processes. The loss of energy through neutrino emission would have the effect of reducing T if energy were not absorbed to make good the loss.

(2) In increasing the energy of rotation of the star.

(3) In increasing the ratio n_N/n_4^2.

As the effect of neutrino emission is more conveniently dealt with at a later stage (Section 8) this sink of energy will be neglected for the moment. In addition the absorption of energy in increasing the energy of rotation only becomes important when the star is nearing the state of rotational instability. Thus, since in the present discussion the star is not regarded as being in any imminent danger of becoming rotationally unstable, the energy absorbed in (2) may also be neglected.

The gravitational potential energy $-\Omega$ of the star is given by

$$-16\pi^2 G \int_0^R \rho(r) r\, dr \int_0^r \rho(r')(r')^2\, dr', \tag{34}$$

where R is the radius of the star. If we assume that whatever the values of M, R the star is always a member of a single homologous series, it follows from (34) that Ω is of the form

$$\Omega(M, R) = BM^2/R, \tag{35}$$

where B is independent of M, R. This assumption would not be permissible in an accurate calculation but can properly be used in making an order of magnitude estimate for Ω. At first sight it might seem that the energy released (in the sense

* Once the point E has been passed the amount of energy radiated from the surface of the star is negligible compared with the energy absorbed in these processes.

of the previous paragraph) in a contraction of the star from R to $R-\delta R$ should be given by $\Omega(M, R-\delta R)-\Omega(M, R)$ but this is not the case because energy has to be supplied merely in order to increase the density of the stellar material. Thus according to (33′) the total "energy of degeneracy" of the star is given by

$$6{\cdot}1 \times 10^{15}\pi \int_0^R (\rho')^{\frac{4}{3}} r^2\, dr.$$

In an order of magnitude calculation the contribution of the neutrons to ρ can be neglected, and we may put $\rho' = \rho$ in (36). It then follows by using the assumption of the homologous series that the energy of degeneracy is of the form

$$CM^{\frac{4}{3}}/R, \tag{37}$$

where C is independent of M, R.

Since (35) and (37) depend differently on M there must be some value of the mass for which these two quantities are equal. According to Chandrasekhar * this occurs when M is equal to the limiting mass $5{\cdot}75\ M_\odot/\mu_e^2 = M_{cr}$ say. Thus for $M > M_{cr}$ we have

$$\frac{\Omega}{\text{Total energy of degeneracy}} \doteq \left(\frac{M}{M_{cr}}\right)^{\frac{4}{3}}.$$

Moreover, since the total energy released is given by

$$\Omega - [\text{Total energy of degeneracy}],$$

it follows that

$$[\text{Total energy released}] = [\text{Total energy of degeneracy}].\ [(M/M_{cr})^{\frac{4}{3}} - 1]. \tag{38}$$

Now the energy of degeneracy *per cm.*^{3} is given by (33′) so that the energy released *per cm.*^{3} can be estimated as

$$1{\cdot}53 \times 10^{15}(\rho')^{\frac{4}{3}}\{(M/M_{cr}^{\frac{4}{3}}) - 1\}.$$

This expression can be replaced by the two equations

$$\left.\begin{aligned} [\text{Energy released per cm.}^3] &= 1{\cdot}53 \times 10^{15}\alpha(\rho')^{\frac{4}{3}}, \\ \alpha &\doteq (M/M_{cr})^{\frac{4}{3}} - 1. \end{aligned}\right\} \tag{39}$$

The purpose of this apparently trivial step is that values of α can be specified and the first of equations (39) can be regarded as an *accurate* equation for the energy released. The second of equations (39) then becomes an equation that determines M *approximately* in terms of the specified α. In other words, instead of specifying M and calculating the energy release approximately we can specify the energy release and calculate the mass approximately. Since it is the energy release that is required in the present work and not the mass the second alternative is preferable. It might seem that the procedure of specifying the energy release renders the above investigation superfluous. This is not the case, however, for we have to be sure that the values we specify are reasonable values and this cannot be decided unless the above discussion of order of magnitude is given.

The energy absorbed in increasing the thermal energy of the stellar material and the energy density of radiation is small compared with (33′). Thus since the energy absorbed in increasing the rotation of the star and in making good the loss of energy by neutrino emission is being neglected in the present discussion,

* S. Chandrasekhar, *op. cit.* p. 424.

effectively the whole of the energy released is to be regarded as absorbed in the production of free neutrons. The energy required to produce one free neutron is about $m_N c^2/125$. Thus the free neutron density is given by dividing the energy released per cm.^{3} by $m_N c^2/125$. This gives

$$\log n_N = 20{\cdot}1 + \frac{4}{3}\log\rho' + \log\alpha. \tag{40}$$

This equation plays a very important part in the following discussion. The value of M is given approximately in terms of α by the second of equations (39). A selection of corresponding values of M and α are given in the following table:—

TABLE IX

M/M_{cr}	1	2	3	5	7	10	15
α	1	0·59	1·08	1·92	2·66	3·64	5·08

If the value of $\log n_N$ given by (40) is substituted in (25) the following equation between T and ρ' is obtained:—

$$T = \frac{4{\cdot}73(7{\cdot}57 - 0{\cdot}271y)}{11{\cdot}03 - 13/12\log\rho' + 9/8\log T - \log\alpha}. \tag{41}$$

Equation (41) gives the track of the stellar material in the helium zone. We now see that the neglect of the rotational energy and the energy lost in neutrino emission has very little effect on the track of the material. Absorption of energy by these processes could be included by suitably altering the value of α in (41), but since α appears only through its logarithm the consequent change in T would be very small.

It is convenient in numerical calculations based on (41) to put $9/8\log T$ equal to a representative constant. For the case $\alpha = 0{\cdot}24$, $M/M_{cr} = 1{\cdot}4$, put $9/8\log T = 1$. The values of T corresponding to a set of values of ρ' are then easily obtained.

TABLE X

$\alpha = 0{\cdot}24$, $M/M_{cr} \simeq 1{\cdot}4$.

ρ' (gm. per cm.^{3})	10^8	10^9	10^{10}
y	3·32	7·5	16·6
T (in units of 10^9 °C.)	7·9	9·0	8·0

The error introduced by putting $\log T = 8/9$ is evidently very slight. Similar sets of values can be obtained for other values of α (that is, for other values of M/M_{cr}). An important result emerges from the values given in Table X. The two points $\rho' = 10^8$, $T = 7{\cdot}9$; $\rho' = 10^9$, $T = 9{\cdot}0$ lie within the helium zone (see 4.1), but the point $\rho' = 10^{10}$, $T = 8{\cdot}0$ lies in the heavy element zone. This means that the track of the stellar material passes through the helium zone *and moves out again into the heavy element zone*. This is a general result and does not depend on the particular value of α used in Table X. Thus if we are to follow the track of the material at densities higher than 10^9 gm. per cm.^{3} it is necessary also to work out the track in the heavy element zone. This question will now be considered.

7. *The Track of Stellar Material in the Heavy Element Zone*

If A, Z are chosen so that $(Q_A^Z - 0{\cdot}543yZ)/A$ is a maximum with respect to *both* A and Z then in the heavy element zone

$$\rho' \simeq A n_A^Z m_P,$$

because, as shown in Section 4, these values of A, Z give the most abundant heavy element. If this maximum value is denoted by $(Q_A^Z/A - 0{\cdot}543yZ/A)_{\max}$ then we obtain from equation (19′)

$$\log n_N = 34{\cdot}08 + \frac{3}{2}\log T - \frac{4{\cdot}73}{T}\left(\frac{Q_A^Z}{A} - \frac{0{\cdot}543yZ}{A}\right)_{\max}, \tag{42}$$

since the last term on the right-hand side of (19′) is negligible when $\rho' \simeq A n_A^Z m_P$ (it will be recalled that for the most abundant element $A > 50$). When $\rho' < 10^9$ gm. per cm.^{3} the contribution of the y term is small and

$$\left(\frac{Q^Z}{A} - \frac{0{\cdot}543yZ}{A}\right)_{\max} \simeq \left(\frac{Q_A^Z}{A}\right)_{\max}.$$

In this case $(Q_A^Z/A)_{\max}$ is determined, as in Section 4, by using the values of Q_A^Z measured in the laboratory. At very high densities, however, the term in y becomes important and it is no longer possible to obtain $(Q_A^Z/A - 0{\cdot}543yZ/A)_{\max}$ directly from laboratory data. It is then necessary to use an expression for Q_A^Z/A as an explicit function of A, Z. Such an expression due to Weizsäcker is available and can be written in the form *

$$\frac{Q_A^Z}{A} = 14{\cdot}9 - 21\frac{(A-2Z)^2}{A^2} - \frac{14{\cdot}2}{A^{\frac{1}{3}}} - \frac{0{\cdot}625Z^2}{A^{\frac{4}{3}}} \tag{43}$$

The first term on the right-hand side is the main term; the second term is a correction arising from the Fermi-Dirac statistics satisfied by the proton and neutron; the third term is a surface effect; the fourth term represents the electrical energy of the nucleus arising from the Coulomb forces between the protons. This expression is in good agreement with the measured values of Q_A^Z/A over the whole range of the periodic table. Using (43) we have

$$\frac{1}{A}(Q_A^Z - 0{\cdot}543yZ) = -(5{\cdot}1 + 14{\cdot}2A^{-\frac{1}{3}}) + \frac{Z}{A}(84 - 0{\cdot}543y) - \frac{Z^2}{A^2}(84 + 0{\cdot}625A^{\frac{1}{3}}). \tag{44}$$

The maximum value of (44) with respect to Z is given by

$$\frac{2Z}{A} = \frac{1 - 0{\cdot}00647y}{1 + 0{\cdot}00744A^{\frac{1}{3}}} \tag{45}$$

and is equal to

$$-(5{\cdot}1 + 14{\cdot}2A^{-\frac{1}{3}}) + \frac{21(1 - 0{\cdot}00647y)^2}{1 + 0{\cdot}00744A^{\frac{1}{3}}}. \tag{46}$$

The maximum value of (46) is given by putting $A = A_{\max}$, where

$$A_{\max} = 45{\cdot}44\,\frac{(1 + 0{\cdot}00744A^{\frac{1}{3}})^2}{(1 - 0{\cdot}00647y)^2}. \tag{47}$$

Substituting $A = A_{\max}$ in (45) and (46) gives

$$\left.\begin{aligned}\left(\frac{Q_A^Z}{A} - \frac{0{\cdot}543yZ}{A}\right)_{\max} &= \frac{954{\cdot}2}{A_{\max}}(1 + 0{\cdot}00744A_{\max}^{\frac{1}{3}}) - (5{\cdot}1 + 14{\cdot}2A_{\max}^{-\frac{1}{3}}) \\ \left(\frac{2Z}{A}\right)_{\max} &= \frac{1 - 0{\cdot}00647y}{1 + 0{\cdot}00744A_{\max}^{\frac{1}{3}}},\end{aligned}\right\} \tag{48}$$

* H. A. Bethe and R. F. Bacher, *Rev. mod. phys.*, **8**, 82, 165, 1936.

where $(2Z/A)_{\max}$ is the value of $(2Z/A)$ for the element of maximum abundance.

It follows from the equations derived above that $A_{\max}$ and $(Q_A^Z/A - 0{\cdot}543yZ/A)_{\max}$ are functions of ρ'. The following table gives a selection of values of $A_{\max}$, $(2Z/A)_{\max}$, and $(Q_A^Z/A - 0{\cdot}543yZ/A)_{\max}$ for various ρ'. (It may be noted that the free neutrons cease to obey "classical" statistics when the density exceeds the values given in this table. The statistical equations would require modification before they could be applied to degenerate neutrons.)

TABLE XI

$\log \rho'$	10·08	11·04	11·27	11·49	11·59	11·79
y	18·4	38·2	45·7	54·1	58·6	68·0
$A_{\max}$	75	110	130	160	180	240
$\left(\frac{Q_A^Z}{A} - \frac{0{\cdot}543yZ}{A}\right)_{\max}$	4·67	2·10	0·84	−0·45	−1·07	−2·27
$\left(\frac{2Z}{A}\right)_{\max}$	0·78	0·65	0·59	0·53	0·50	0·44

The above values show that for sufficiently high ρ' the most abundant element in the heavy element zone lies at the top end of the periodic table. This is one of the most important results derived in the present paper. It arises directly from the term in y which is due to the degeneracy of the electrons. The values given for $(2Z/A)_{\max}$ show that at these high densities the ratio of neutrons to protons present in the heavy elements is far higher than is found in the laboratory.

The discussion of the track of the stellar material in the heavy element zone of the T, ρ'-diagram is similar to that given in 6.2 for the helium zone, except that in place of equation (25) for the helium zone we now have equations (42), (47) and (48). These latter equations together with (40) determine the track of the material in the heavy element zone. The equation for T corresponding to (41) is

$$T = 4{\cdot}73 \frac{\{(954{\cdot}2/A_{\max}).(1 + 0{\cdot}00744A_{\max}^{\frac{2}{3}}) - (5{\cdot}1 + 14{\cdot}2A_{\max}^{-\frac{1}{3}})\}}{13{\cdot}98 + 3/2\log T - 4/3\log\rho' - \log\alpha}, \qquad (49)$$

where $A_{\max}$ is given in terms of y (and consequently in terms of ρ') by (47). Thus (49) gives T as a function of ρ' and therefore gives the required track of the stellar material in the T, ρ'-diagram. It is again convenient to put $\log T$ equal to a representative constant in the numerical work. The following table gives values of T, n_N, n_P for a set of values of ρ' for the case $\alpha = 0{\cdot}24$:—

TABLE XII

	$\alpha = 0{\cdot}24$,	$\log T = 0{\cdot}85$.		
$\log \rho'$	10·08	10·56	11·04	11·27
T	9·09	10·78	8·64	4·73
$\log n_N$	32·9	33·6	34·2	34·5
$\log n_P$	27·7	27·3	22·8	9·7

The outstanding feature of this table is the very rapid decrease in n_P when ρ' rises to 10^{11} gm. per cm.^{3}. The reason why n_P decreases so rapidly is that equations (40) and (42) prevent T rising appreciably above 10 (if T were to rise appreciably above 10 then equation (42) would give a value of n_N that was too large to be consistent with (40). This limitation in temperature is a direct result of the absorption in free neutron production of the energy released in the collapse of the star). Thus as ρ' increases the second term on the right-hand

side of (18) increases until it becomes comparable with $\log n_N$. At this stage both n_P and T decrease very rapidly. Thus beyond this stage it is probably unsafe to employ the equations of statistical equilibrium. This means that the relative abundances of the elements can no longer be calculated on a statistical basis. The problem evidently becomes of extreme complexity, for nuclear reactions can continue to be important for some elements and can become unimportant for others. Indeed it seems possible that some elements may become frozen simply as a result of a sufficiently large increase in density.

The values given in Tables X and XII depend on the value $\alpha = 0{\cdot}24$ used in the computations. For other values of α a curve similar to that shown in Fig. 1 can be obtained. In particular the values of T, ρ' at the point where statistical equilibrium breaks down depend on α. The values of T, ρ' near this break-down point are given in the following table for a selection of values of α:—

TABLE XIII

α	1·07	2·69	3·89	5·49
T (at break-down point)	9·79	11·58	12·50	14·55
ρ' (at break-down point)	11·27	11·49	11·59	11·79
Atomic weight of most abundant element at break-down point	130	160	180	240

The values given in Table XIII show the very important result that for sufficiently large α the most abundant element at the break-down point lies at the upper end of the periodic table.

In following the track of the material in the T, ρ'-diagram we have referred explicitly to the material at the centre of the star. The above discussion enables the case of material not at the centre to be immediately included, for the only difference that can arise is through a difference in the value of the parameter α. It is to be expected that α will have a different value in different parts of the star (α could only be constant throughout the star if there was a very special distribution of material in the star. There seems to be no reason why a rapidly collapsing star should in general satisfy such a special distribution). This means that there will be a variation in both the statistical break-down point and the element of maximum abundance at the break-down point as we pass from one part of the star to another.

The discussion of the track of the stellar material given in the present and previous sections depends on equation (40). This equation applies only so long as insufficient energy is released by the contraction of the star to convert effectively the whole of the material into neutrons. If, however, the collapse of the star continues after material reaches the break-down point, a stage must eventually be reached at which equation (40) no longer holds. At this stage energy is again available to raise the temperature of the material and this has the effect of increasing n_P. Thus it is possible for statistical equilibrium to be resumed if the contraction of the star should proceed sufficiently far before rotational instability occurs. It is unnecessary for the purpose of the present

paper to consider the properties of the material in this new state, and this will not be attempted.

8. *The Chemical Composition of Material thrown off by certain Rotationally Unstable Collapsing Stars*

8.1. *Preliminary Remarks.*—In the present section the results so far obtained will be used to give a discussion of the relative abundances of the elements distributed in interstellar space by the process of rotational instability. Since we are considering a star with mass appreciably greater than Chandrasekhar's limit it follows that collapse must continue until rotational instability occurs.* For the sake of definiteness a star of initial mass 20 $M_\odot$ will be considered, so that a large fraction of the mass must be thrown off in the instability process in order that the mass of the remaining nucleus be reduced to a value comparable with Chandrasekhar's limit.

When the star becomes rotationally unstable the material thrown off expands against the powerful gravitational field of the star. At first sight the expansion seems to be exactly the reverse of the compression that occurred during the contraction of the star, but it turns out that there is an important difference between these two processes. If we consider the stellar material during the stage of compression *before* the break-down point is reached, and we imagine that some cause momentarily reduces the temperature sufficiently for nuclear reactions effectively to cease, then absorption of energy by the nuclear reactions will also effectively cease. Thus a surplus of energy will be available that raises the temperature of the material until statistical equilibrium is again set up. This means that before the break-down point is reached the material *during compression* is automatically governed to remain in statistical equilibrium. An opposite situation occurs during the expansion of the material thrown off by rotational instability. For if nuclear reactions cease during expansion then the supply of energy *decreases* since energy is no longer made available by the nuclear reactions. Accordingly there is no automatic governing effect during expansion and material once frozen will become even more effectively frozen as the expansion proceeds.

In view of the remarks of the previous paragraph it is evidently plausible to suppose that some of the ejected material remains frozen during the expansion, and in the following discussion it will be assumed that the quantity of this frozen material amounts to a few per cent. of the total material thrown off by the instability process. Thus the present discussion divides into two parts:—

Case 1.—To obtain the composition of material that moves during the expansion in a reverse direction along a track in the T, ρ'-diagram similar to that shown in Fig. 1.

Case 2.—To obtain the composition of material that remains frozen during the instability process.

The quantity of material satisfying Case 2 is only a few per cent. of the material satisfying Case 1. Case 1 is considered in 8.2 and Case 2 in 8.3.

8.2. *The composition of material in Case* 1.—The final composition of this material is determined by the composition at the time it crosses the line AB, since the material must freeze when T decreases below 4×10^9 °C. Thus the final

* F. Hoyle. In course of publication.

composition is close to that occurring at the point D in the T, ρ'-diagram. It will be shown that the composition at this point is roughly uniform over the lower half of the periodic table.

The relative abundance $n_{A'}^{Z'}/n_A^Z$ is easily obtained from equation (19) in terms of y, T. Thus we have

$$\log n_A^Z = 34{\cdot}08 + \frac{3}{2}\log T + \frac{3}{2}\log A + \frac{A}{A'}\left(\log n_{A'}^{Z'} - 34{\cdot}08 - \frac{3}{2}\log T - \frac{3}{2}\log A'\right) + \frac{4{\cdot}73A}{T}\left(\frac{Q_A^Z}{A} - \frac{Q_{A'}^{Z'}}{A'} + 0{\cdot}543y\left\{\frac{Z'}{A'} - \frac{Z}{A}\right\}\right). \quad (50)$$

On the line AB of Fig. 1 we have $T=4$, and to a sufficient approximation the experimental values of Q_A^Z/A can be used. In addition the term in y can be neglected when ρ' is of order 10^7 gm. per cm.^{3}. To this approximation equation (50) does not involve ρ', and if $n_{A'}^{Z'}$ is specified for a given pair of values A', Z' the corresponding values of n_A^Z can be worked out for all other A, Z. The density ρ is then given by

$$\rho \simeq \rho' \simeq m_P \sum_{A,Z} A n_A^Z,$$

since the contribution of the free neutrons to the density of the material is small when $T=4$, $\rho' = 10^7$ gm. per cm.^{3}. Thus specifying $n_{A'}^{Z'}$ is equivalent to specifying ρ'. In fact it can be shown by calculation that if we specify n_{56}^{26} for example, then ρ' is roughly proportional to n_{56}^{26} for values of ρ' lying between A and B in Fig. 1. The value $n_{56}^{26} = 10^{28}$ was adopted in calculating the following selection of values of n_A^Z (it being more convenient to specify $n_{A'}^{Z'}$ than to specify ρ):—

TABLE XIV

Element	He^4	O^{16}	Si^{28}	Fe^{56}	Cu^{63}	Kr^{82}	Sn^{118}	Pb^{208}
$\log n_A^Z$	27·2	20·2	23·7	28	26·7	18·8	−3·3	−222·8

The total density given by this distribution is close to 10^7 gm. per cm.^{3}. Thus the distribution at the point D in Fig. 1 has approximately the form given in Table XIV.

The rather small oxygen abundance given in Table XIV arises from adopting an approximation in which the whole of the material freezes at $T=4$. In a more accurate treatment it would be necessary to take account of the fact that nuclear reactions among light elements will persist to a lower temperature than for nuclei with $A>30$. This must be the case because the proton and α-particle energies required to penetrate the Coulomb barriers is less for the light nuclei. Thus in the cooling of the material a temperature will be reached at which nuclear reactions can be regarded as taking place for $A<30$, but as being effectively frozen for $A>30$. When this occurs the large helium content shown in Table XIV tends to be converted into elements with atomic weights up to 30. This effect can be seen by considering the helium-oxygen ratio at temperatures less than $T=4$. From equation (50) we then have

$$\log n_{16}^8 = 34{\cdot}08 + \frac{3}{2}\log T + \frac{3}{2}\log 16 + 4(\log n_4^2 - 34{\cdot}08 - \frac{3}{2}\log 4\mathrm{T}) + \frac{4{\cdot}73}{T}(Q_{16}^8 - 4Q_4^2). \quad (51)$$

Putting $Q_{16}^8/16 = 8{\cdot}54$, $Q_4^2/4 = 7{\cdot}57$ it is easy to see from (51) that n_{16}^8 becomes

comparable in magnitude with n_4^2 when T is close to 3. For this temperature it can be shown that

$$\log n_N \simeq \log n_P \simeq 21.$$

It may be concluded that if equilibrium among light nuclei persists down to $T=3$ then the helium present at $T=4$ will be largely converted into elements lying between carbon and silicon.

8.3. *The composition of material in Case 2.*—The results given in Table XI show that the composition of material that remains frozen during the expansion depends on the density attained at the onset of the instability process. The most abundant element present in a particular sample of this material must consequently vary according to the region of the star that supplies the material in question. Thus, if for the sake of definiteness, the central density is taken to be 5×10^{11} gm. per cm.^{3} then material derived from near the centre of the star must have $A_{\max}$ close to 240, while on the other hand material derived from the outer parts of the star will only attain a density comparable with the mean density of the star. Allowing for the central density exceeding the mean by a factor of 20, it follows that the latter material attains a density of order $2{\cdot}5 \times 10^{10}$ gm. per cm.^{3}, and according to Table XI such material must have $A_{\max} < 110$. It is to be expected therefore that the average composition of all material satisfying Case 2 will consist of a fairly uniform distribution of the elements over the upper half of the periodic table, since such an average must include contributions from all parts of the star.

It is tacitly assumed in the previous paragraph that material can attain a density of 5×10^{11} gm. per cm.^{3} without the statistical break-down point being reached. The values given in Table XIII show that this will be the case if the parameter α is greater than 5.

The discussion of the composition in the case where material reaches the break-down point is more complicated than that given above, since it is then necessary to replace the statistical equations by a consideration of individual reactions. This discussion will not be attempted in the present paper. It may however be noted that the slowing down of nuclear reactions at the break-down point must make such material particularly likely to remain frozen during the subsequent expansion.

9. *The Effect of β-processes on the Rate of Collapse of a Star*

9.1. *The Rate of Collapse of the Star.*—It is convenient at this stage to discuss the condition (iv) of Section 5. The requirement is that a sufficient number of interchanges between neutrons and protons must take place for equation (18) to be applicable.

If the time required for (18) to be established is less than the time of free fall of the star, then the star will collapse freely along the portion FG of the track of stellar material in the ρ', T-diagram. On the other hand, if the time of free fall is less than the time required for (18) to be established then the evolution of the stellar material along FG is determined by the time required for (18) to be established and not by the time of free fall. The reason for this is clear, for if equation (18) is not satisfied, free neutrons will not be produced at the rate calculated above. But it is the production of free neutrons that removes energy and is responsible for the rapid collapse of the star. This means that free fall of

the star is not possible unless equation (18) is satisfied. The situation is that the star will collapse at a rate given either by the time required for condition (iv) to be established or by the time of free fall, according to which of these two times is the longer. The purpose of the present section is to estimate these two times.

The time required for the free fall of the star is easily estimated, so far as an order of magnitude is concerned, from the formula $\sqrt{(R^3/GM)}$ where R is the radius of the star at F. This formula can be written as $\sqrt{(3/4\pi G\bar{\rho})}$ where $\bar{\rho}$ is the mean density of the star at the time when the material near the centre is at F. This requires a central density of about 2×10^8 gm. per cm.^{3} which corresponds to a mean density of about 10^7 gm. per cm.^{3}. The value of $\sqrt{(3/4\pi G\bar{\rho})}$ for $\bar{\rho}=10^7$ gm. per cm.^{3} is 0·42 sec., which may be compared with the result obtained below, that the time required for (18) to be established is of the order of 100 sec. It is seen therefore that the rate of collapse is determined by the time required for equilibrium to be set up between the protons and neutrons.

9.2. *Concerning β-processes.*—It is seen from Table XI that $(2Z/A_{\max})$ is close to 0·5 when ρ' exceeds 10^{11} gm. per cm.^{3}. This value is much less than the value of $2Z/A$ occurring under laboratory conditions. Thus even for uranium the quantity $2Z/A$ is as large as 0·77 while for lighter elements $2Z/A$ is of course greater still. It follows therefore that the nuclei considered in the discussion of Section 7 would be extremely unstable against β-disintegration under laboratory conditions. Thus such nuclei thrown off by a rotationally unstable star must rapidly be converted into nuclei with the values of $2Z/A$ found on Earth. An estimate for the period of decay under laboratory conditions can easily be given. The upper limit of the β-spectrum is given by putting $2Z/A=0{\cdot}5$ in $(Q_A^Z-Q_A^{Z-1})$. The value of $(Q_A^Z-Q_A^{Z-1})$ is given to sufficient accuracy by differentiating Q_A^Z with respect to Z. This gives

$$84\left(1-\frac{2Z}{A}\right)-\frac{1{\cdot}25Z}{A^{\frac{1}{3}}}\text{ m.m.U.,} \tag{52}$$

and on putting $2Z/A=0{\cdot}5$ we obtain an upper limit to the β-spectrum of $(42-0{\cdot}312/A^{\frac{1}{3}})$ m.m.U., which is close to 39 M.V. Now according to the Fermi theory of β-decay the lifetime of an allowed transition is given by

$$\frac{2^5\pi^4g^2W_0^5}{15h^7c^6}, \tag{53}$$

when the following assumptions are made:—

(i) The upper limit W_0 of the spectrum is appreciably greater than the rest mass of the electron.

(ii) The effect of the Coulomb field of the nucleus on the wave function of the emitted electron is neglected.

(iii) The matrix element for the β-transition is put equal to unity, so that g is the β decay constant.

The requirement (i) is satisfied by a large margin in the present problem, but (ii) and (iii) are not satisfied for nuclei at the top end of the periodic table. However, it is known that the effect of the Coulomb field is to decrease the lifetime of an unstable nucleus while on the other hand the matrix element contributed by the initial and final wave functions of the nucleus increases the lifetime as

the atomic weight A increases. The laboratory data shows that these two effects approximately cancel each other, so that (53) may be used as an approximate formula over the whole range of atomic weight.

The constant g is not known from theoretical considerations and the usual procedure is to fix g from the observed lifetime of a typical β-decaying substance. As a reasonable estimate a lifetime of one second for an allowed β-reaction with a spectrum of upper limit 7·5 M.V. may be taken. Thus the case discussed above in which the upper limit is 39 M.V. would have a lifetime of $2{\cdot}6 \times 10^{-4}$ sec. It is particularly to be noted that this lifetime refers to laboratory conditions and could not arise for material in a star. The reason for this is that the emission of electrons is prevented by degeneracy. Thus if we take a nucleus with a value of $Z > Z_{\max,}$ electrons cannot be emitted because the effect of degeneracy is to fill the electron levels to a value *higher than the upper limits of the β-spectra.* When $Z = Z_{\max}$ the upper limit of the β-spectrum becomes equal to the top of the energy levels filled by the electrons while for $Z > Z_{\max}$ the β-spectra have upper limits that exceed the highest energies of the degenerate electrons. This situation was of course taken care of in Section 7 when the maximum value of $(Q_A^Z - 0{\cdot}543yZ)$ with respect to Z was obtained.*

It follows from the remarks of the previous paragraphs that under equilibrium conditions β-disintegrations in a star only occur for nuclei with $Z < Z_{\max.}$ Thus in order to obtain the rate at which β-disintegrations occur under statistical equilibrium it is necessary to estimate the ratio $n_A^Z / n_A^{Z\max}$ for $Z < Z_{\max.}$ This ratio is obtained immediately by putting $A' = A$, $Z' = Z_{\max}$ in equation (50), which gives

$$\log\left(\frac{n_A^Z}{n_A^{Z\max}}\right) = -\frac{4{\cdot}73}{T}\{(Q_A^{Z\max} - 0{\cdot}543yZ_{\max}) - (Q_A^Z - 0{\cdot}543yZ)\}, \qquad (54)$$

Expanding the right-hand side of (54) by Taylor's theorem gives

$$\log\left(\frac{n_A^Z}{n_A^{Z\max}}\right) = -\frac{2{\cdot}36}{T}\cdot(Z_{\max} - Z)^2\left\{\frac{d^2}{dZ^2}(Q_A^Z - 0{\cdot}543yZ)\right\}_{Z=Z_{\max},}$$

$$\text{when } 0 < (1 - Z/Z_{\max})^2 \ll 1, \text{ since } \left\{\frac{d}{dZ}(Q_A^Z - 0{\cdot}543yZ)\right\}_{Z=Z_{\max}} = 0. \qquad (55)$$

Equation (55) gives the required formula for the number of nuclei that are unstable against β-decay.

Now the expression (53) can no longer be used on account of electron degeneracy. For a nucleus of charge Z such that $0 < 1 - Z/Z_{\max} \ll 1$ the difference between the upper limit W_0 of the β-spectrum and the highest energies of the degenerate electrons is given by

$$\frac{(Z - Z_{\max})^2}{2}\left\{\frac{d^2}{dZ^2}(Q_A^Z - 0{\cdot}543yZ)\right\}_{Z=Z_{\max}} = \frac{84(Z - Z_{\max})^2}{A}\Big\}. \qquad (56)$$

Thus β-disintegration can only occur if the energy W of the emitted electron satisfies the inequality

$$W_0 > W > W_0 - \frac{84(Z - Z_{\max})^2}{A}, \qquad (57)$$

* At first sight it seems strange that the mathematics deals so easily with this matter. But this is the outstanding feature of statistical mechanics. So long as statistical equilibrium occurs the equilibrium properties of an assembly can be determined without direct reference to individual processes being necessary.

where W, W_0 are measured in milli-mass units. The fraction of the electrons emitted in this range is given by

$$\int_{W_0-84(Z-Z_{\max})^2/A}^{W_0} W(W^2/m_e^2c^4-1)^{\frac{1}{2}}(W_0-W)^2dW \div \int_0^{W_0} W(W^2/m_e^2c^4-1)^{\frac{1}{2}}(W_0-W)^2dW, \quad (58)$$

where m_ec^2 is the rest mass of the electron, measured in m.m.U. The result (58) follows immediately from the well-known spectrum

$$W(W^2/m_e^2c^4-1)^{\frac{1}{2}}(W_0-W)^2dW$$

given by the Fermi theory. When $W_0/m_ec^2 \gg 1$ and $84(Z-Z_{\max})^2/A \ll W_0$, the expression (58) is given to sufficient accuracy by

$$10\left\{\frac{84(Z-Z_{\max})^2}{W_0A}\right\}^3. \quad (59)$$

Thus the lifetime against β-disintegration in a star is given by dividing (53) by (59). Using the value $2{\cdot}6\times10^{-4}$ sec. obtained above for (53) this gives a lifetime of

$$2{\cdot}6\times10^{-5}\left\{\frac{W_0A}{84(Z-Z_{\max})^2}\right\}^3 \text{ sec.}$$

Since it is sufficiently accurate to put $W_0=42$ m.m.U. for all $Z<Z_{\max}$ and

$$\frac{n_A^Z}{n_{A\max}^Z} = 10^{-\frac{4{\cdot}73}{T}\cdot\frac{84(Z-Z_{\max})^2}{A}}$$

it follows that the number of β-disintegrations per sec. given by all nuclei of atomic weight A is

$$\frac{3{\cdot}1\times10^5 n_A^{Z_{\max}}}{A^3}\sum_{Z<Z_{\max}}(Z-Z_{\max})^6 10^{-\frac{4{\cdot}73}{T}\cdot\frac{84(Z-Z_{\max})^2}{A}} \quad (60)$$

and the *total* number of β-disintegrations occurring per cm.^{3} per sec. is

$$3{\cdot}1\times10^5\sum_A\frac{n_A^{Z_{\max}}}{A^3}\sum_{Z<Z_{\max}}(Z-Z_{\max})^6 10^{-\frac{4{\cdot}73}{T}\cdot\frac{84(Z-Z_{\max})^2}{A}} \quad (61)$$

The main contribution to the series in (60) comes from Z close to $Z_{\max}$ for all values of A. In order to evaluate (60) for various A it is necessary to specify $4{\cdot}73/T$. Over the portion of the T, ρ'-diagram that is of interest (that is, between F and G) we may take 0·5 as a representative value for $4{\cdot}73/T$. The following table gives expression (60) divided by $n_A^{Z_{\max}}$, for a set of values of A:—

TABLE XV

A	16	56	126	252
Expression (60)/$n_A^{Z_{\max}}$	0·18	0·43	0·65	0·92

Table XV shows the surprising result that the number of β-disintegrations is a roughly constant multiple of $n_A^{Z_{\max}}$ for all values of A from 56 up to the top of the periodic table. It follows therefore that (61) is given approximately by

$$\frac{2}{3}\sum_A n_A^{Z_{\max}} \simeq \frac{2\rho'}{3m_PA_{\max}}.$$

Thus we reach the important conclusion that the number of β-disintegrations occurring per cm.^{3} per sec. for material in the heavy element zone is of order $2\rho'/3m_P A_{\max}$ no matter what value of ρ' is chosen on the portion F to G of the track of the stellar material.

Under equilibrium conditions the number of inverse β-processes involving electron capture must be equal to the number of direct disintegrations calculated above. Thus the number of inverse β-process is of order $2\rho'/3m_P A_{\max}$ per cm.^{3} per sec. A direct calculation of the number of inverse β-processes could be given but it is more convenient to follow the method given above. It may be noted that the inverse β-processes arise from electron capture by nuclei with $Z > Z_{\max}$. The reason for this is that the energy of the captured electrons must exceed the upper limits of the β-spectra and this requires $Z > Z_{\max}$. This result deserves comment for it means that the direct β-disintegrations occur for $Z < Z_{\max}$ while the inverse processes occur for $Z > Z_{\max}$. Thus there can be no detailed balancing. The *total* number of direct β-disintegrations in an equilibrium state is of course equal to the number of inverse processes even if there is no detailed balancing in individual reactions. At first sight this lack of detailed balancing seems to be in opposition to the well-known argument of statistical mechanics that seeks to establish detailed balancing from a consideration of the Hermitian character of the Hamiltonian of the assembly. But there is no such contradiction because the argument based on the Hermitian property of the Hamiltonian depends on the assembly being in strict thermodynamic equilibrium, whereas our assembly is not in thermodynamic equilibrium because the density of free neutrinos is effectively zero (the system is in equilibrium in the sense that average values can be calculated for n_A^Z, n_P, n_N and n_ϵ, but this does not constitute strict thermodynamic equilibrium and is really a steady state distribution). The breakdown of thermodynamic equilibrium is seen as soon as we consider reactions involving neutrinos. Thus a neutrino is always *emitted* in a state of positive energy, but a neutrino can never be *captured* in a state of positive energy. Accordingly the neutrino reactions are not thermodynamically reversible and it is therefore not surprising that detailed balancing does not occur in these reactions.* It is precisely this property of the neutrino reactions that requires the direct β-disintegrations to occur for $Z < Z_{\max}$ and the inverse reactions to occur for $Z > Z_{\max}$.

We are now in a position to estimate the length of time required for statistical equilibrium between neutrons and protons to be set up. The way in which statistical equilibrium comes to be set up is easily understood. Consider a departure from statistical equilibrium in which

$$\left.\begin{aligned} &n_A^Z > (n_A^Z)_{\text{equilibrium}}, \quad \text{when} \quad Z > Z_{\max}, \\ &n_A^Z < (n_A^Z)_{\text{equilibrium}}, \quad \text{when} \quad Z < Z_{\max}, \\ &\Sigma n_A^Z = \sum_A (n_A^Z)_{\text{equilibrium}}, \end{aligned}\right\} \tag{62}$$

This has the effect of reducing the number of direct β-disintegrations below the equilibrium value, while the rate of the inverse β-processes on the other hand

* Gamow and Schoenberg, *Phys. Rev.*, **59**, 539, 1941, use detailed balancing in neutrino reactions in order to calculate n_A^{Z-1}/n_A^Z. The writer is not in agreement with this assumption as the value of n_A^{Z-1}/n_A^Z must be calculated from the full scheme of equations involving equilibrium between the nuclei and cannot be calculated from the neutrino reactions alone.

is increased. This follows immediately since the direct disintegrations occur for nuclei with $Z<Z_{\max}$ and the inverse processes for $Z>Z_{\max}$. Thus under the conditions (62) the rate of the inverse processes must exceed the rate of the direct disintegrations. Moreover the rate of the direct processes can be reduced to a negligible value by putting $\sum_{Z<Z_{\max}} n_A^Z \simeq 0$. The rate of the inverse processes on the other hand cannot be much increased above the equilibrium value. Accordingly it follows that an excess of inverse processes over direct disintegrations amounting to $1/2\ m_P$ per gm. can be produced in a time of order

$$\frac{1}{2m_P}\cdot\frac{2}{3A_{\max}m_P}=\frac{3A_{\max}}{4}\text{ sec.},$$

which is of the order of 100 sec.

The above investigation shows that the time taken for β-processes to establish statistical equilibrium between protons and neutrons is appreciably longer than the time of free fall of the star which is of order one second (see 9.1). Thus the time scale of the contraction is determined by the β-processes and not by the time of free fall. This result enables the loss of energy by neutrino emission to be calculated fairly accurately, since the total number of β-processes occurring before the onset of rotational instability cannot appreciably exceed $1/2m_P$ per gm. Now the great majority of these reactions occur for nuclei with Z close to $Z_{\max}$ and consequently the energy of emission of the neutrino must be small compared with the upper limits of the β-spectra. Expressed quantitatively the energy of an emitted neutrino must be less than

$$\frac{84(Z_{\max}-Z)^2}{A}\text{ m.m.U.}$$

Thus it follows, since $A_{\max}>50$, that 1 M.V. is a representative estimate for the average energy of the emitted neutrinos. Accordingly the energy lost by neutrino emission cannot be much greater than 5×10^{17} ergs per gram of stellar material. This value may be compared with the energy required to produce the large number of free neutrons present in material with density greater than 10^{11} gm. per cm.^{3}. The energy required to liberate these neutrons from the nuclei is approximately $c^2/135$ per gm. Accordingly the energy required to liberate the neutrons is about 7×10^{18} ergs per gm. as against 5×10^{17} ergs per gm. for the loss of energy by neutrino emission. This result shows that the energy lost in neutrino emission is small compared with the energy absorbed by the other nuclear reactions. This provides justification for our neglect of the neutrino energy loss in the work of Sections 6 and 7.

Finally the important point may be noted that because the time-scale for the contraction of the star is determined by the time required for equilibrium to be set up in the β-processes, the material of the star is never strictly in a steady equilibrium state. In a fuller discussion it would be necessary to take account of variations from the steady state and this would undoubtedly involve a more complicated theory than that given above. This has not been attempted in the present paper because it is desirable to begin by exploring the equilibrium theory. Moreover the present method of determining the track of stellar material in the T, ρ'-diagram already takes account of the main features associated with the changing state of the material.

10. *Novae, Supernovae and the Origin of the Solar System*

The work of previous sections shows that when material near the centre of a collapsing star reaches the point E of Fig. 1 the hydrostatic equation of support ceases to hold even as an approximation. Thus the collapse and the onset of rotational instability is catastrophic *provided rotational instability does not occur before the point E is reached.* Collapsing stars that satisfy the latter condition will be regarded as supernovae.*

According to this theory the outburst of a supernova is even more rapid than is suggested by observation, which gives a time of about two days for the rapid rise to maximum light. The explanation of the slower rise of the observed light is clear, for the supernova will not emit appreciable energy in the visible part of the spectrum until the material thrown off has expanded to an envelope of dimensions large compared with the radius of the star. As the envelope expands high frequency radiation is degenerated and the proportion of light emitted in the visible spectrum increases. Thus the time taken for the light emitted in the visible spectrum to reach its maximum is associated with the time required for the expansion of the envelope and is not directly connected with the time-scale of the instability process (provided this time-scale is less than the time required for the expansion of the envelope).

If a star becomes rotationally unstable before the point E in Fig. 1 is reached then the equation of support (1) will be satisfied to a good approximation throughout the collapse. This means that the state of rotational instability will be approached gradually over a time that is of the order of, or greater than, a thousand years. It seems likely that instability in this case will take the form of a series of minor explosions rather than the star remaining completely stable until a state of such violent instability is reached that an outburst of the supernova type occurs. That is, the star will splutter as it slowly contracts. Thus the point E divides collapsing stars into two groups. The first group of stars reach E and undergo a catastrophic outburst, whilst stars not reaching E undergo a series of minor explosions similar in character to the outbursts of an ordinary nova.

The existence of the point E in the T, ρ-diagram provides a natural theory of the difference between novae and supernovae. Moreover in this theory there is no invention of an unspecified physical process in order to produce the energy required for the material to be thrown off, since this energy is derived from the gravitational potential energy of the star. In comparing the theory with observation it is therefore necessary that the magnitude of the gravitational potential energy of the remaining stellar nucleus after outburst shall be greater than the kinetic energy at infinity of the material thrown off. If $M\mu_e^2/5{\cdot}75M_\odot \gg 1$ the material thrown off has a mass comparable with M, and the remaining stellar nucleus has a mass comparable with Chandrasekhar's limit $5{\cdot}75M_\odot/\mu_e^2 = M_{cr}$. Thus if v is the velocity at infinity of the ejected material it is necessary that $Mv^2/2$ shall be of the same order but less than GM_{cr}^2/R. It is of interest to apply this condition in the case of the supernova of 1054 A.D. The observational data lead to the values $M \simeq 15M_\odot$, $v \simeq 10^8$ cm. per sec., $R = 10^9$ cm. Thus since $5{\cdot}75M_\odot/\mu_e^2 \simeq 1{\cdot}5M_\odot$, it is seen that $Mv^2/2 \simeq 1{\cdot}5 \times 10^{50}$ ergs while $GM_{cr}^2/R \simeq 6 \times 10^{50}$ ergs. Accordingly the supernova of 1054 A.D. satisfies the requirements of the theory of rotational instability.

* This may be compared with the Gamow-Schoenberg neutrino emission theory.

The distribution of the elements in the solar system raises a question of interest in relation to the present work. It is easy to see that the theory of the synthesis of the elements given in Section 8, taken together with the theory of the origin of the solar system recently proposed * by the writer, provides a satisfactory explanation of this distribution so far as elements heavier than helium are concerned. It is important to notice that this original planetary material contains no hydrogen. The present hydrogen content of the solar system is regarded as due to the sweeping up of interstellar hydrogen by the envelope of planetary material. In order to consider this process quantitatively let s be the radius of the original envelope of planetary material, t the time required for the condensation of the gaseous material into solid bodies, u the velocity of the Sun relative to the interstellar hydrogen, and σ the density of the interstellar hydrogen. Then the amount of hydrogen acquired by the planetary system is of order

$$\pi s^2 \sigma u t.$$

A recent investigation by the writer suggests that t must be of order 3×10^8 years. With this value for t, and $s = 3 \times 10^{14}$ cm., $\sigma = 10^{-22}$ gm. per cm.^{3}, $u = 2 \times 10^6$ cm. per sec., we obtain a mass of hydrogen amounting to about twenty per cent. of the total mass of the planets. This is of the required order of magnitude.

St John's College,
Cambridge:
[1946 April 5.

* F. Hoyle, *Proc. Camb. phil. Soc.*, **40**, 256, 1944; *M.N.*, **105**, 175, 1945.

NOTE ON THE ORIGIN OF COSMIC RAYS

F. Hoyle

(Received 1946 April 6)

Summary

The question of associating the origin of cosmic rays with properties of highly collapsed stars is discussed. It is shown that the observed intensity of cosmic rays can be explained by means of such an association. The general principles involved in the origin of cosmic rays are briefly considered.

---

1. It has been suggested that cosmic rays may originate in supernova outbursts.* The question arises as to whether any of the processes discussed in the preceding paper provide support for this suggestion. The values given there in Table I, for the case of a star of mass $10M_\odot$ and an initial rotational velocity before collapse of 1 km. per sec., would appear to provide a starting-point for such a theory. The values for this case show that instability will not occur until the rotational velocity of the star is of the same order as the velocity of light. Now it is to be expected that in the process of rotational instability the velocity with which the material is thrown off to infinity will be of the same order, but less than, the rotational velocity just before instability occurs. Thus cases may occur in which material is thrown off from a rotationally unstable star with a velocity of the same order as the velocity of light C. The kinetic energy possessed by a proton would then be of order $m_PC^2 = 9{\cdot}3 \times 10^8$ e.V., where m_P is the mass of the proton. That is, the kinetic energy is of the same order as the rest-mass.

It is important to notice that m_PC^2 is the *average* kinetic energy of emitted protons. The discussion given below in Section 3 shows how particles can acquire energies that are considerably in excess of this average value. Indeed it will be seen that 10^9 e.V. is the basic unit out of which the cosmic rays of very high energy are built. In this connection it may be noted that the observed cosmic ray energy spectrum depends on E according to the factor dE/E^3. This shows, for example, that the combined energy of all cosmic rays with energies greater than 10^{15} e.V. is less than the combined energy of all cosmic rays with energies greater than 10^{10} e.V. by a factor 10^{-10}. Thus the fraction of the total energy in the form of cosmic rays of very high energy ($>10^{15}$ e.V.) is extremely small. It follows therefore that the discussion of Section 3 which attempts to explain the origin of these cosmic rays is concerned with very exceptional cases rather than with the average situation.

2. For the above process to occur, the potential GM/R at the surface of the star at the onset of rotational instability must be of order C^2. This condition gives

$$\bar{\rho} \simeq 3C^6/4\pi G^3M^2 \simeq 10^{17}(M_\odot/M)^2 \text{ gm. per cm.}^3,$$

where the mean density $\bar{\rho}$ is defined by $4\pi\bar{\rho}R^3/3 = M$. Thus provided $M < 30M_\odot$,

* W. Baade and F. Zwicky, *Phys. Rev.*, **45**, 138, 1934 and **46**, 76, 1934 ; F. Cernushi, *Phys. Rev.*, **56**, 120, 1939.

the mean density in the star attains a value of order 10^{14} gm. per cm.^{3} before the process described in Section 1 occurs. Now at such densities the average distance apart of the particles becomes comparable with the range of nuclear forces. At this stage the whole star passes over into effectively one "atomic nucleus". This nucleus must be almost entirely composed of neutrons, and is similar to the "neutron cores" of Baade and Zwicky.* This result can be understood from the work of the previous paper, which shows that increasing density is characterized by increasing neutron content, and that already at a density of 10^{12} gm. per cm.^{3} material must be largely composed of neutrons.

It is important to notice that the cohesive effect of general nuclear forces, existing throughout a star in which $\bar{\rho} > 10^{14}$ gm. per cm.^{3}, cannot prevent rotational instability occurring. The reason for this is firstly that owing to the saturation property of nuclear forces the binding energy per neutron in stellar material of density 10^{14} gm. per cm.^{3} will not be much different from the value occurring in ordinary atomic nuclei. This gives a binding energy due to nuclear forces that is less than $10^{-2}C^2$ per gram of material, which is only about one per cent. of the sum of the rotational and gravitational energy of the material. Moreover the nature of nuclear forces is such that the nuclear binding energy cannot be significantly increased by compressing the material to densities higher than 10^{14} gm. per cm.^{3}.

3. The remarks of the preceding sections indicate that cases occur in which material with density exceeding 10^{14} gm. per cm.^{3} is thrown off by rotationally unstable stars. As a result of nuclear transformations and of internal collisions such material will be gradually broken down into protons, neutrons and ordinary laboratory nuclei. But before reaching this final state, the material must pass through an intermediate stage in which it consists largely of nuclear lumps with the following properties:—

$$A \gg 10^6, \quad Z/A \ll 1,$$

where Z and $A-Z$ are the numbers of protons and neutrons respectively in the nuclear lumps. It seems possible that a "thunder-storm effect" may occur during this intermediate stage. Thus consider an assembly that consists initially of positively charged nuclear lumps together with a sufficient density of free electrons to preserve electrical neutrality. Then suppose that the nuclear lumps become separated from the free electrons by a process in which the energy required to produce the separation is supplied at the expense of the kinetic energy of the nuclear lumps. On this hypothesis it is clear that large electric potentials must be built up within the distribution of nuclear lumps. The order of magnitude of the largest potential that can arise is given by

$$ZeV = WA,$$

where V is the potential attained (taking the potential at infinity as zero) for the special case in which the nuclear lumps all contain Z protons, $A-Z$ neutrons, and all have an initial kinetic energy WA.

Consider now the effect of a nuclear transformation resulting in the emission of a proton from one of the nuclear lumps, and let the transformation occur *after* the charge separation has taken place. The proton must then be repelled by the powerful electric field arising from the whole distribution of positively

* *Loc. cit.*

charged nuclear lumps, and will reach infinity with a kinetic energy of order eV. That is, the proton is repelled to infinity with energy WA/Z. Thus, since $A/Z \gg 1$, it follows that such a proton acquires an energy that is large compared with W. Now according to Section 1 the quantity W is of order $m_P C^2$, so that the present process results in protons acquiring energies that are large compared with $m_P C^2$. This is what is required to explain the origin of the cosmic rays of very high energy.

It is important to notice that on the present theory the cosmic rays may contain laboratory nuclei as well as protons. For it may well be the case that the nuclear lumps emit these nuclei in addition to protons, and if such nuclei are able to reach infinity without suffering complete break-up due to collisions, then it is to be expected that they will form a component of the cosmic rays.

The above discussion is incomplete in the important respect that the charge separation is not discussed in detail. It would be difficult to give such a discussion, because the motion of material during the process of rotational instability is extremely complicated. For this reason no attempt will be made to specify a definite process. It is of interest, however, to note that charge separation can occur either through the effect of magnetic fields or by motion through a medium that is highly resistant to free electrons, but which allows the nuclear lumps to pass freely. Finally it may be remarked that the conditions described above are only required for the cosmic rays of very high energy. As pointed out in Section 1 these cosmic rays carry a very small fraction of the total energy. Thus the process of charge separation need only be of rare occurrence.

4. The problem of the origin of the cosmic rays is, in its nature, outside the range of laboratory process. In these circumstances the most plausible type of theory is an extrapolation from a process already observed to occur in a less extreme form. The process of rotational instability suggested above fulfils this condition since unstable stars are observed where material is thrown off with velocities of several thousand km. per sec. Thus the difference between the suggested mode of origin of the cosmic rays and the observed process is one of degree rather than of principle. As was seen in the preceding paper the question of degree depends on the initial rotation of the star before collapse begins. If a star with mass appreciably greater than Chandrasekhar's limit has a sufficiently small initial rotation it seems probable that the required process will occur. Suitably small initial rotations might well occur in about one collapsing star in ten. On the basis of one supernova per nebula per 500 years, the frequency of collapsing stars giving cosmic ray energies may reasonably be taken as one per nebula per 5000 years.

In view of the desirability of obtaining confirmation of the suggestions made above it is fortunate that a calculation can be made, on the basis of the above remarks, for the mean energy density in space of the cosmic rays. As pointed out by Compton and Chou*, if the mean energy density taken throughout space is put equal to the observed energy density in the neighbourhood of the earth, then the total energy of all the cosmic rays in the universe comes out to be considerably larger than the total energy radiated by all the stars in the universe in a time of 10^{10} years. An explanation of this remarkable result must evidently be regarded as crucial to any proposed theory of the origin of cosmic rays. This

* A. H. Compton and P. Y. Chou, *Phys. Rev.*, **51**, 1025, 1937.

question has already been remarked on by Baade and Zwicky* but in view of its great importance it seems worth discussing the matter at greater length than is given by these authors.

Consider two concentric spheres of radii r and $r+dr$, the centre of the spheres being at the point where it is desired to calculate the intensity of the cosmic rays. The number of nebulae lying between the two spheres can be taken as $4\pi\nu r^2 . dr$, where ν is the mean space density of extra-galactic nebulae. The cosmic ray energy produced by these nebulae is of order

$$4\pi\nu r^2 dr(M_\odot C^2/\tau) \text{ per sec.},$$

where τ is the rate per nebula at which cosmic ray stars occur, and where the energy produced by each such star is taken as $M_\odot C^2$ (it will be remembered that these stars all have mass greater than Chandrasekhar's limit). The intensity of cosmic rays at the centre due to the spherical shell is given immediately on dividing this expression by $4\pi r^2$ and is accordingly $\nu M_\odot C^2 dr/\tau$. The total cosmic ray intensity is then obtained by integrating the latter expression from $r=0$ to $r=R$, where R is the radius of the universe. This gives an intensity

$$\nu M_\odot C^2 R/\tau \text{ per unit area per unit time.}$$

Putting $R=3\cdot10^8$ parsecs, which is close to the observed radius of the universe (the values of the radius of the universe indicated by theoretical studies are not much in excess of $3\cdot10^8$ parsecs), $\nu=3\cdot10^{-73}$ corresponding to a mean internebular distance of $5\cdot10^5$ parsecs and $\tau=$ one cosmic ray producing star per nebula per 5000 years, gives an intensity of 10^{-3} ergs per cm.^{2} per sec. This value is in astonishingly good agreement with the observed value of the cosmic ray intensity for the neighbourhood of the earth.

The above theory relates the following physical quantities:—(i) the cosmic ray intensity, (ii) the radius of the universe, (iii) the mean internebular distance, (iv) the frequency of supernovae, by means of a formula involving C and $M_\odot$. These quantities are of such a diverse nature that this formula could hardly represent a tautology. Thus the present formula would not appear in a theory that did not involve collapsing stars. It appears, therefore, that the remarkable agreement between the calculated and the observed cosmic ray intensities must be regarded as lending very powerful support to the suggestion that the origin of cosmic rays is associated with the properties of collapsing stars (the presence of such a large factor as 10^{-73} in the above formula makes it very unlikely that the agreement is sheer coincidence).

5. The formula obtained in Section 4 for the cosmic ray intensity is based on the tacit assumption that the point in space at which the intensity is calculated is not very close to a cosmic ray producing star. It is of interest to estimate how near such a star the point may be taken before the contribution from the star rises to a value equal to that calculated above. Let d be the distance from a star that produces a cosmic ray energy $M_\odot C^2$. If v_1, v_2 are the greatest and least velocities respectively (relative to the point in question) of the cosmic rays produced, then the cosmic rays will reach the point at a time d/vs after the outburst, and will continue to pass the point until a time d/v_1 after the outburst.† It follows that

* W. Baade and F. Zwicky, *loc. cit.*

† The outburst itself lasts only a few seconds, as was shown in the previous paper. Thus for the present purpose the outburst can be regarded as occurring at a definite instant.

the flood of cosmic rays will last for a time $d(v_1-v_2)/v_1v_2$ and that the mean flux of cosmic ray energy from the star during this time is of order

$$\frac{v_1v_2M_\odot C^2}{4\pi d^3(v_1-v_2)} \text{ per unit area per unit time.}$$

As an example put $v_1-v_2=10^{-2}C$, $v_1v_2\simeq C^2$, which gives an intensity of

$$\frac{10^2C^3M_\odot}{4\pi d^3} \text{ per unit area per unit time.}$$

This is equal to the value 10^{-3} ergs per cm.^{2} per sec. obtained in Section 4 when $d\simeq 7\cdot 10^{22}$ cm. This distance is considerably less than the average distance between the nebulae. It follows therefore that if the cosmic ray intensity due to the star is to rise above the general background produced by the universe as a whole, then the star and the point in question must lie within the same nebula, or in very close neighbours.

The results obtained above may be applied to the Galaxy. Suppose that a cosmic ray producing star occurs at a distance of 10^{22} cm. from the Earth. Then the intensity from the star would become greater than the general background and moreover the angular distribution would become non-isotropic. This effect would last for a time

$$10^{22}(v_1-v_2)/v_1v_2\simeq 100 \text{ years.}$$

A comparison of this time with a frequency of one cosmic ray producing star per nebula per 5000 years shows that it is only during comparatively brief periods that the cosmic ray intensity can rise above the general background level discussed in Section 4. This background level remains practically constant over a time interval of order R/C which is about $3\cdot 10^9$ years when we put $R=3\cdot 10^8$ parsecs, and moreover this background level is isotropic, provided the distribution of nebulae at large distances is isotropic.

6. The preceding work concerns one particular process of origin of the cosmic rays. The important question arises as to whether other processes along similar lines can be expected to occur. In particular the idea of a cosmic "thunderstorm effect" would seem worthy of a more general examination. The essential features of the thunderstorm effect are:—

(1) The assembly must contain massive charged particles that are of such a nature that their ratio of charge to mass is extremely small compared with e/m_P. Moreover these particles must possess charge of the same sign, and the remainder of the assembly must contain ordinary ionized gas with a sufficient charge bias to compensate for the charge of the massive particles, thereby giving electrical neutrality to the complete assembly.

(2) A process of charge separation occurs that removes the ionized gas and leaves the massive particles. The massive particles must remain charged throughout this process.

(3) An elementary particle (or ordinary atomic nucleus) with charge of the same sign as the charge on the massive particles must be produced within the assembly of massive particles after the charge separation process has taken place. Furthermore, conditions must be such that this particle can move without resistance, thereby enabling the particle to be repelled freely to infinity by the combined electric field of the massive particles.

For the reasons given in Section 4 it is believed that the main energy of cosmic rays arises in processes associated with highly collapsed stars. It is possible, however, that subsidiary effects may occur in many ways. In particular it is possible that subsidiary processes may make an important contribution to the cosmic rays of very high energy, since these cosmic rays carry only a small fraction of the total cosmic ray energy. As an illustrative example we may consider the case of large interstellar dust particles which must become negatively charged if they are situated in a region containing ionized gas.* Under such conditions we have an assembly that satisfies requirement (1). The requirement (2) is also satisfied if it is assumed that the dust particles maintain their charge throughout an encounter between this assembly and a cloud of comparatively dense *neutral* gas. For if the density of the neutral gas is suitably adjusted the effect of such an encounter is to sweep out the original ionized atoms without appreciably affecting the dust particles. Moreover it can be shown that if the density of the neutral gas is sufficiently high this charge separation process can take place without "lightning flashes" occurring. The requirement (3) cannot be satisfied so long as the dust lies within the neutral gas, but at the end of the encounter when the dust passes out of the neutral gas it is possible for conditions to occur that satisfy this requirement. Thus an electron breaking away from a dust particle at the end of the encounter would be repelled to infinity by the field of the dust particles, and it is easy to show that such an electron can acquire an energy of order 10^{13} e.V. without requiring an unduly high density for the large dust particles.

The example given in the previous paragraph involves such simple concepts as to suggest that the principles involved in the origin of cosmic rays may well be very straightforward. Indeed even if some point should on closer investigation be found to render this particular process untenable, the above remarks have significance in that they show the essential simplicity of the "thunder-storm effect" whereas this simplicity is somewhat obscured in the discussion of Section 3 owing to such unusual circumstances as rotational instability and the occurrence of nuclear lumps.

St. John's College,
Cambridge:
1946 *April* 5.

* L. Spitzer, *Ap. J.*, **93**, 369, 1941.

ERRORS OF TIME DETERMINATION

L. S. T. Symms and C. Joy A. Penny

(Communicated by The Astronomer Royal and The Astronomer Royal for Scotland)

(Received 1946 May 25)

Summary

The investigation refers to clock errors determined with a (non-reversible) transit circle of standard type, the collimation errors being measured against collimators and the level errors obtained by nadir observations. It is shown that there is a strong correlation between the clock errors and the measured collimation errors indicating that a good deal of the day-to-day scatter in clock error is due to accidental error in the measurement of collimation. Apart from this a good deal of the personal equation in clock error for an observer, that is, the (slowly changing) systematic error in his determinations of time, is found to arise from personal equation in his measures of collimation. The effect of errors of nadir setting in the measurement of level error is also discussed. For the instrument investigated (the Edinburgh Transit Circle) it is found that the probable error of a single night's determination of time (using the mean of ten stars) is ± 32 milliseconds when no corrections for personal equation are applied. By applying personal equations this figure is reduced to about ± 20 milliseconds, and of this that part of the probable error not due to errors in measuring collimation and level is ± 10 milliseconds or less. This figure applies to a hand-driven micrometer and includes contributions arising from erroneous star places.

---

1. For the last five years observations of clock error have been made with the Transit Circle at the Royal Observatory, Edinburgh to provide, in combination with those obtained at the Abinger station of the Royal Observatory, Greenwich, the basis of the Greenwich Time Service.

The introduction of quartz clocks has, among other things, made it easier to derive the probable error of a single night's determination of clock error. The erratics of rate of quartz clocks are very small, the mean absolute scatter of daily rate being of the order of $\pm 0{\cdot}05$ milliseconds per day, and this means that over a considerable period of time the plot of clock error against date is of the form $E = a + bt + ct^2$ where E is the clock error and t is the time in days reckoned from some convenient zero date. (The term ct^2 arises from the frequency drift of the quartz oscillator – the frequency changes at a steady rate after the initial "ageing" has been completed.) Accordingly, given a set of measured clock errors extending over a period of months, a quadratic form can be fitted to them by a least square adjustment, and the probable error of a single night's determination of time then can be deduced from the residuals in the ordinary way. But previous to the introduction of quartz clocks the pendulum clocks in use were subject to erratics of relatively large amount—so large that it was necessary to credit the clock with a natural tendency to wander and so to fit a

curve, often of a sinusoidal form, to the transit observations. This was mathematically equivalent to introducing a larger number of unknowns into the equations of condition used to determine the clock error as a function of date and the uncertainty in the p.e. of a night's determination of clock error was thus enhanced.

By referring the transit observations to quartz clocks it appears that the p.e. of a time determination (10 stars) at Edinburgh is of the order of ± 30 milliseconds. But the internal agreement of the transits of the individual stars used for a single night's determination indicates a probable error much smaller than this, and it has long been suspected that the measurement of instrumental corrections—collimation and level errors—contributes very heavily to the total p.e. The present paper is an attempt to determine these contributions on a numerical basis. The investigation arose from a remark by Professor Greaves that there appeared to be some evidence that a systematic difference in the measures of collimation error by the two observers (LS and JP) remaining at Edinburgh was responsible for an apparent large difference of personal equation between the clock errors obtained by them. This suspicion was confirmed, as will be seen, but the centre of interest shifted as the investigation proceeded; and it became apparent that not only were differences in personal equation in the measurement of collimation largely responsible for correspondingly large systematic differences in clock error between the various observers—as was only to be expected—but that there was a striking correlation between the residuals of the collimation measures from the means for the various observers and the corresponding residuals in clock error. In other words, quite apart from the systematic differences to which the term "personal equation" is usually applied, the day-to-day scatter in measurements of clock error is largely due to the day-to-day scatter in measurements of collimation with this instrument, i.e. the probable error of the clock errors, is largely due to the probable error of the measures of collimation and the probable error of measurement of the actual transits is relatively small. This is a striking result which, as far as we know, has not been demonstrated before on a numerical basis.

2. The transit circle is of standard type and incorporates a hand-driven impersonal micrometer. Its aperture is 8·5 inches and its focal length is 8 feet 11 inches. It is a reversible instrument, but over the period discussed it was not reversed. The collimation errors are determined against two collimators, and the level error by nadir settings with a Bohnenberger eyepiece in the normal way. A defect in the instrument is the employment of collimators with apertures less than the aperture of the transit instrument itself; the collimator apertures are 6·1 inches, their focal length being 6 feet 3 inches. The collimators are mounted in the same observing house as the main instrument; their mountings are satisfactorily stable and the seeing conditions for the collimator operations are very good. Care is taken to achieve steady and uniform illumination of the collimator wires. A single wire is mounted in the focal plane of each collimator and the north collimator is set on the south prior to setting the instrument on each. Over the period discussed a reversing prism was *not* employed in any of the measurements. The transit instrument is carefully focussed for the R.A.s, the travelling wire being in the focal plane of the objective.

No attempt was made to determine time fundamentally, i.e. to determine star places, and to relate them to the Sun. For each night's work some ten stars were observed and the transit times, after correction for collimation and level, were subtracted from the FK3 Right Ascensions as published in *Apparent*

Places of Fundamental Stars. The resulting differences were plotted against the tangent of the declination and a straight line fitted to the plotted points. The clock error was then deduced from the ordinate of the point of this line corresponding to abscissa $\tan\phi$, where ϕ is the latitude ($=55^\circ\,55'{\cdot}5$ for Edinburgh). The azimuth error corresponds to the slope of the line, but the stars were chosen by the observer so that the mean $\tan\delta$ approximates closely to $\tan\phi(=1{\cdot}48)$. This ensures that the resulting clock error is practically independent of the azimuth error of the instrument. It is stressed that for each night's work the transits were reduced using the collimation and level errors as measured by the observer on duty on the night concerned. Corrections for adopted personal equations for two of the observers were applied to the clock errors: +18 milliseconds for LS and +3 milliseconds for B.

The transits were observed against a sidereal free pendulum clock (Shortt No. 4) and this clock was compared with the quartz clock A1 (a mean time clock) before, during and after each period of observation so that the result of each night's work was referred at once to the quartz clock and gave the error of the latter. Further, since the Rugby G.B.R. radio time-signals at 10^h U.T. were recorded against A1, the transits can be transferred to any other clock, situated at any station, against which the 10^h G.B.R. radio time-signals were also recorded, since daily comparisons between such a clock and A1 are given by the radio link. In the present investigation the transits have been used, in this way, to give the clock errors of the quartz clock IVC at the Post Office Radio Research Laboratories, Dollis Hill. The Dollis Hill clocks were more suitable than the Observatory quartz clock since, being battery controlled, they are not subject to dislocations due to mains' surges (the clock A1 at Edinburgh has been mains operated). Of the three clocks available at Dollis Hill, IVC was selected as being the clock which had put up the best performance over the period considered. Accordingly, all transits have been transferred so as to give, for each occasion, a clock error of IVC (a *positive* error denoting clock *slow*) and these clock errors have, in effect, been plotted against date. But it is desirable that over a lengthy period the plotted points should lie approximately on a straight line when a chart with a large scale of ordinates is used and that the first differences of the plotted points should be reasonably small. Accordingly, a quadratic expression in t (the term in t^2 representing the frequency drift of the oscillator) is removed from the figures before plotting, where t is the time in days from a chosen epoch. The resulting plot for IVC is shown in Fig. 1 (Plate 11). Each observer's transit points are joined by straight lines, not as an indication of the performance of the clock, but merely to assist the eye in separating the observations of the different observers. Other features of the figure will be described later.

3. If on any particular occasion the measured collimation is in excess of its true value by an amount c then (since in determining the level error the reading for the nadir setting is combined with the measured reading for the line of no collimation) the level error will also exceed its true value by c (in addition to the effect of an error in the nadir reading) and the final effect of the error c on clock error is that the determined clock error is in excess of its correct value by $-2c\sec\phi$, in addition to the mean error of observation of the transits themselves and the errors originating from an erroneous nadir setting. Accordingly, the clock errors will be correlated with the collimation errors owing to the common errors c and the correlation coefficient will indicate the extent to which erroneous

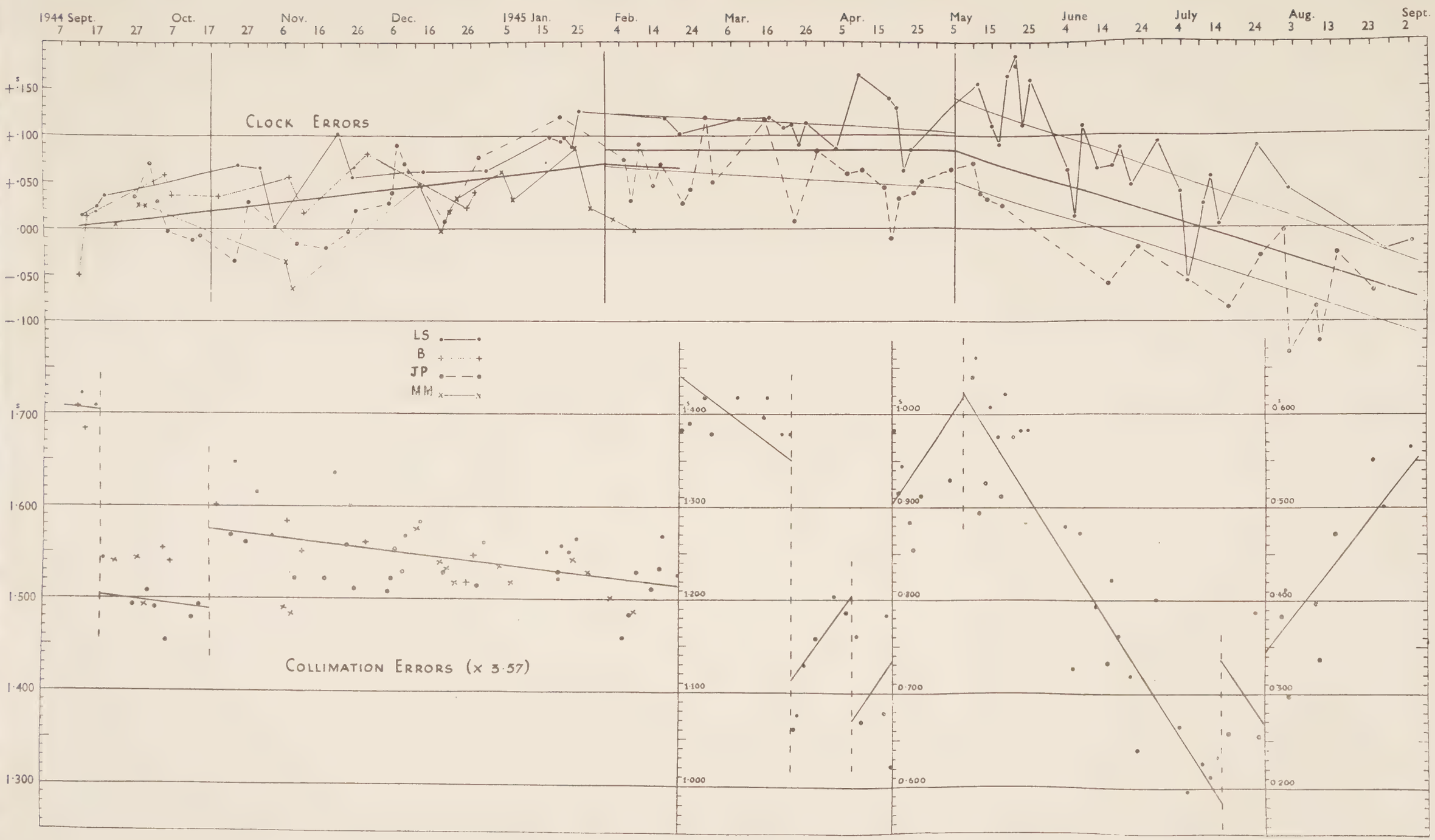


FIG. 1.

L. S. T. Symms and C. J. A. Penny, Errors of Time Determination.

measures of collimation contribute to the total probable error of time determination. As regards the effect of erroneous nadir settings on the clock errors if l be the excess of the measured level error over its true value owing to this cause, the resulting excess in clock error will be $-l \sec \phi$; thus there should be correlation between the clock errors and the nadir readings and the correlation coefficient should indicate the extent to which the errors l contribute to the total probable error of the clock errors.

The following general remarks are made at this point:—

(*a*) We shall be concerned with correlations between quantities which are themselves subject to secular change. Clock errors can clearly change progressively with date owing to clock rate and it is found that the collimation errors and nadir readings of the Edinburgh Transit Circle are also subject to secular drift. Now we are concerned with the calculations of correlation coefficients between the *accidental* parts of the quantities involved. Clearly then we must analyse, not the quantities themselves but their residuals from datum lines which take up the secular drifts. Thus instead of clock errors we must consider the residuals from a datum line drawn on the plot of clock error against date. Similarly for the collimation errors and nadir readings.

(*b*) We have already emphasized that we are investigating the correlation between the day-to-day variations of the clock errors and the instrumental errors. This means that we cannot use the same datum lines for all observers. Each observer is subject to personal equations in his transit measures and in his measures of the instrumental errors, and we are concerned with the accidental scatter of the measures about values representing the personal equations. Further, we have noticed that the personal equations are subject to slow secular changes. Accordingly, since we are dealing with the correlations between the accidental departures from the slowly changing values corresponding to the personal equations, we must aim at taking up the personal equations and their secular changes in the datum lines. It thus becomes necessary to adopt a different set of datum lines for each observer, and we shall be concerned with the evaluation of correlation coefficients between the residuals of the individual measures from these lines. The methods used for obtaining the datum lines will be described below; they incidentally give information regarding personal equations, but we shall not be primarily concerned with these.

4. Fig. 1 (Plate 11) shows that over the whole period considered there were two small changes of rate of IVC, towards the end of January and towards the beginning of May. The transit observations are not accurate enough to define the dates where these changes occurred, but fortunately other data were available. The three Group IV clocks at Dollis Hill are compared in pairs, the oscillations of the crystals being made to produce beats by virtue of the slight departures of the actual frequencies from the nominal value of 100 kilocycles per second, and the beats are counted by electrically actuated counters. The daily totals of the counters give the relative errors of the three clocks in units of one-hundredth of a millisecond, and these totals have been supplied by the Post Office Radio Branch. From them it appears that the changes of rate took place around January 31 and May 5 but were apparently spread over a few days. These two dates have been adopted as the dates where the changes of rate occurred, and the changes of rate were evaluated from the clock errors of the observers LS and JP, means of the changes of rate given by these two observers being adopted.

The clock errors over the short period January 31 to February 20 were then corrected for the effect of the change of rate on January 31, and over the period October 19 to February 20 least square solutions of the form $y = a + bt$ were fitted to the clock errors for each observer so corrected. As will be seen in Section 5, it was necessary to restrict the investigation to the period October 19 to February 20 because of the large discontinuities in the collimation error. These least square solutions provided the datum lines (shown in Fig. 2) from which clock error residuals were reckoned.

5. The collimation errors appear in the transit reduction forms in equatorial seconds of time. They have been multiplied by $-2 \sec \phi = -3{\cdot}57$ and are plotted in Fig. 1 (Plate 11) beneath the clock errors. At this stage an examination was made to detect any correlation between the two graphs. Throughout the whole period there are some small compact groups of points, usually by the same observer, for which the day-to-day variations of the individual collimation errors are rather strikingly reproduced in the clock errors. There is also evidence that the systematic differences of the measures of collimation by different observers is also reproduced in the clock errors. For the last two-thirds of the whole period this relationship of the personal equations is less obvious, and on close examination it was realized that this is due to sudden discontinuities in the collimation error. On the assumption that the collimation error of the telescope suffered some nine abrupt changes and varied linearly between the changes, preliminary datum lines were drawn from which to measure residuals. These are shown as full lines in Fig. 1 (Plate 11). It still remained to adopt different datum lines for each observer.

It was decided to confine the arithmetical analysis to the period October 19 to Feburary 20, for which the collimation error did not appear to have undergone any abrupt changes. For each observer a datum line was determined by a least square fit of the collimation errors (multiplied by $-2 \sec \phi$) to an expression $a + bt$, where t is the time (in units of 100 days) from a chosen epoch. These datum lines are shown in the lower part of Fig. 2. The residuals of the individual collimation errors from these lines were formed and were available for correlation analysis against the corresponding clock error residuals.

6. In passing we may note that the values of a and b, obtained in the solutions for the clock error and collimation datum lines, give information about the (slowly changing) personal equations. They are collected in the following tables ; the values given are referred to the mean of the four observers; consequently the separate b's do not coincide with the slopes shown in Fig. 2. The epoch, $t = 0$, corresponds to December 20, i. e. to the middle of the period.

Personal Equations. (Values of a.)

Observer	No. of obs.	Clock Error	Collimation
LS	15	+37 msecs.	+38 msecs.
B	6	− 3	− 3
JP	18	− 1	−15
MM	12	−32	−20

These give a correlation coefficient of $0{\cdot}92 \pm {\cdot}10$. The clock error residuals include the provisional personal equations of +18 for LS and +3 for B, previously mentioned. If these are removed and the uncorrected clock errors are used, the correlation is weakened to $0{\cdot}84 \pm {\cdot}19$.

The slopes of the lines, given by the values of b, with the mean removed, are:

Observer	Clock Error	Collimation
	(msecs. per 100 days)	
LS	+ 9	−24
B	−61	−44
JP	+26	+ 9
MM	+26	+59

There appears to be correlation between these which, expressed in the form of a correlation coefficient, gives $0{\cdot}76 \pm {\cdot}25$, indicating that even the different

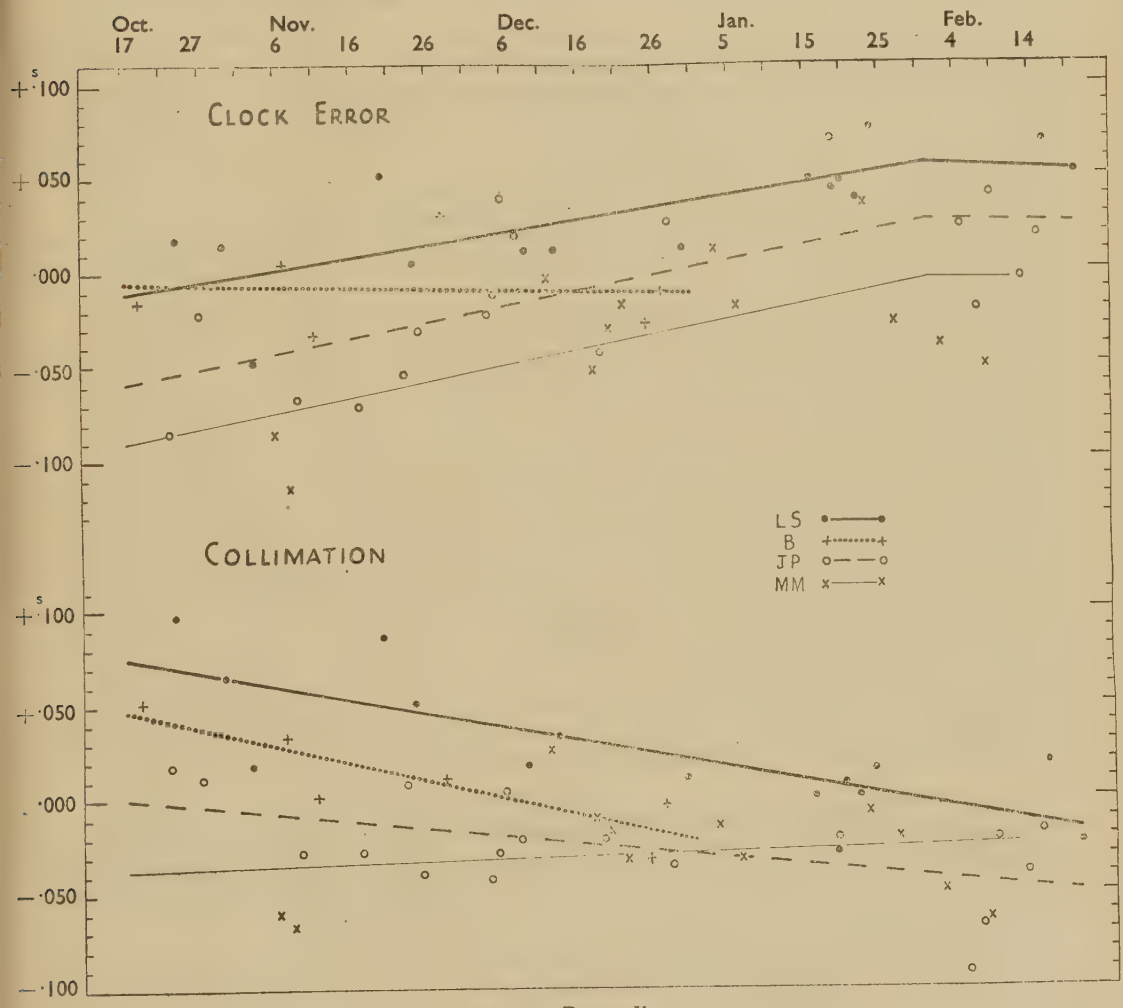


FIG. 2.—*Datum lines.*

slopes in the collimation datum lines are reflected in the clock error residuals, as they should be, when the collimation measures are erroneous.

It should be pointed out that if the real personal equations in clock error are required the effect of these differing ordinates and slopes in the collimation residuals must be eliminated.

7. Returning to our main purpose, we now have two sets of residuals, viz. clock error residuals and collimation residuals, and the coefficient of correlation between them will indicate the extent to which the day-to-day accidental errors of measurement of collimation contribute to the total scatter in the observed clock errors.

Before giving the numerical results we shall investigate the significance of this correlation coefficient in some detail.

Let c = error of measurement of collimation error, a positive value of c meaning that the collimation error is measured too large.

Let C = real excess of collimation error over its mean value.

Let l = error of measurement of level error due to error of measurement of the nadir setting.

Let L = real excess of level error over its mean value.

These quantities are supposed to be expressed in equatorial seconds of time. The mean values referred to are, in each case, functions of some known parameter or parameters, such as the date and temperature.

In obtaining the level error the nadir is combined with the determination of the line of collimation and so the total error of measurement $= l + c$.

The total excess of the *measured* collimation error over its mean value is then

$$\chi = C + c,$$

and the total excess of the *measured* level error over its mean value is

$$\lambda = L + l + c.$$

The real clock error will be a function of date. The clock error actually obtained on any occasion will differ from the real clock error as a result of (i) the error of measurement of the transits themselves, (ii) the error of measurement of collimation and (iii) the error of measurement of level.

As regards (i), let e = that part of the excess of the clock error over its true value due to errors of measurements of the transits.

As regards (ii), the erroneous collimation c produces an erroneous excess $-c \sec \phi$ for the clock error.

As regards (iii), the erroneous level error, $l + c$, produces an erroneous excess $-(l + c) \sec \phi$ for the clock error.

Hence the total excess of the measured clock error over its true value is

$$\epsilon = e - (2c + l) \sec \phi.$$

Consider now a sample of n observations of clock, collimation and level errors. To begin with, let us consider the case in which n is very large and the errors ϵ, χ and λ can be identified with the clock error, collimation and level residuals.

We shall, as usual, denote the standard deviation (S.D.) of any error x by σ_x, so that

$$\sigma_x^2 = \frac{1}{n} \Sigma x^2.$$

We shall also denote by σ_{xy} the mean product (M.P.) of any two correlated errors x, y, so that

$$\sigma_{xy} = \frac{1}{n} \Sigma xy.$$

Let r_1 be the coefficient of correlation between the clock error residuals, ϵ, and the collimation residuals, $-2\chi \sec\phi$.

Hence by the definition of correlation coefficient,

$$r_1 = \frac{-\sigma_{\epsilon\chi}}{\sigma_\epsilon \sigma_\chi},$$

but

$$\sigma_{\epsilon\chi} = \frac{1}{n}\Sigma[e-(2c+l)\sec\phi][C+c]$$

$$= -2\frac{1}{n}\sec\phi \,.\, \Sigma\, c^2,$$

the other sums of products vanishing, assuming that there is no correlation between their components; i. e.

$$\sigma_{\epsilon\chi} = -2\sigma_c^2 \sec\phi, \tag{1}$$

so that

$$r_1 = \frac{2\sigma_c^2 \sec\phi}{\sigma_\epsilon \sigma_\chi},$$

where

$$\sigma_\epsilon^2 = \sigma_e^2 + (4\sigma_c^2 + \sigma_l^2)\sec^2\phi, \tag{2}$$

and

$$\sigma_\chi^2 = \sigma_C^2 + \sigma_c^2. \tag{3}$$

These equations show that a large value of r_1 (i. e. near to unity) will be obtained if σ_c is predominant over the other standard deviations σ_C, σ_e, σ_l.

Up to now we have, for simplicity, considered the number in the sample to be large. We can now drop this restriction provided that, in forming the standard deviations and mean products, we divide the sums by $n-m$ instead of n, where m is the number of unknowns that have been (similarly) introduced into each of the sets of data to define the datum lines. If we treat the residuals for all observers together we shall have $n=51$ and $m=8$ since, for each observer, a linear datum line involving two unknowns was taken out. For the results of an individual observer $m=2$. We ignore the circumstances that the datum lines for the nadir readings (considered later) contain a further unknown, the temperature coefficient, which was determined from all observations over the period considered; and that the datum lines in the clock errors involve a term in t^2, the coefficient of which was determined from a much more extended set of data. A rigorous attempt to take these into account would be very difficult, but the correct results should be intermediate between those obtained by taking $m=8$ and $m=9$, and to the order of accuracy with which we are concerned the error introduced is immaterial. We therefore compute the quantities using $m=8$.

The numerical results for the correlation between the clock error residuals, ϵ, and the collimation residuals, $-2\chi\sec\phi$, for the separate observers and for the mean of all four observers are:

Observer	No. of obs.	r_1 and probable error of r_1
LS	15	$0{\cdot}82\pm0{\cdot}07$
B	6	$0{\cdot}59\pm0{\cdot}27$
JP	18	$0{\cdot}29\pm0{\cdot}16$
MM	12	$0{\cdot}80\pm0{\cdot}09$
All Observers	51	$0{\cdot}59\pm0{\cdot}07$

The probable errors of the correlation coefficients have been calculated from the approximate formula,

$$\text{P.E. of } r = 0{\cdot}6745\frac{1-r^2}{\sqrt{n-m}}\left[1+\frac{11r^2}{4(n-m)}\right],$$

where n is the number of pairs of residuals and m is the number of unknowns introduced into the determinations of each of the datum lines from which the residuals are calculated.*

It is difficult to say dogmatically that the values of r_1 differ for the different observers; but some confirmation is forthcoming, for the two observers LS and JP, from a discussion of the period May 10 to July 14, over which the collimation does not seem to have suffered any abrupt change. During this period LS secured observations on 19 nights and JP on 6, and the values obtained for r_1 over this period are:

Observer	r_1 and probable error of r_1
LS	$0{\cdot}75 \pm 0{\cdot}08$
JP	$0{\cdot}38 \pm 0{\cdot}32$

a result which, for what it is worth, is in agreement with the values for those observers obtained above.

The difference between these two extremes may be partly due to sampling error.† It is not explained by any large difference in the scatter of the correlated quantities for, as will be seen from a table of the collected results given later in Section 11, the probable errors of these quantities are very similar for these two observers. It may be due to an unsuspected correlation between the two components of a mean product which we have assumed to vanish in the expression for r_1. For instance, a bisection error ‡, variable from night to night, in all settings comprising the collimation and level measures (including the preliminary setting of the north collimator on to the south) would introduce a positive correlation between the collimation and level error residuals, and terms containing both c and l would not vanish. The expression for r_1 with these terms retained is:

$$r_1 = \frac{(2\sigma_c^2 + \sigma_{cl})\sec\phi}{\sqrt{[\sigma_e^2 + (4\sigma_c^2 + \sigma_l^2)\sec^2\phi + 4\sigma_{cl}\sec^2\phi]}\sqrt{[\sigma_C^2 + \sigma_c^2]}}.$$

However, the strong correlation between the clock error residuals and the collimation residuals, for most observers, indicates that the errors of measurement of collimations are making a material contribution to the probable error of the time determinations. We can, in fact, obtain an upper limit to that part of the probable error of a time determination not due to erroneous measures of collimation by assuming that the whole of the collimation residuals arise from errors of measurement, i. e. assuming that there are no real fluctuations about the datum lines. On this assumption we take, for each observer, the datum line

* A proof for the case $m=1$ is given by Soper, *Biometrika*, IX, 1913.

† The standard error of sampling for σ_x^2 is $\frac{\sigma_x^2\sqrt{2}}{\sqrt{n-m}}$. For a mean product σ_{xy} it is $\frac{\sigma_{xy}}{\sqrt{n-m}} \cdot \frac{\sqrt{1+r^2}}{r}$. (Professor Greaves has proved this for $m=1$ and Dr Aitken has proved it for the general case.)

‡ In connection with this question of a possible bisection error, a considerable amount of work was done (after the period under discussion) using a reversing prism on all collimation and level settings. Only two observers (LS and JP) were available for this investigation, but fortunately, the personal equations for these two are very different. The results showed that the large difference in personal equation in the measurement of collimation is not merely a question of bisection error ; i. e. although the scatter in the measures was slightly reduced when a reversing prism was used, the difference between the observers' measures was, for some obscure reason, more than doubled.

values to represent the value of his collimation errors and we can thus subtract the collimation residuals from the clock error residuals (i. e. in effect recalculate the clock error with the required collimation error). The probable error of the resulting figures is then that part of the probable error of a determination of time not due to errors of measurement of collimations. The results are as follows:—

Observer	p.e. of clock errors (msecs.)	p.e. of clock errors corrected for collimation (msecs.)
LS	±16	± 9
B	±18	±14
JP	±22	±23
MM	±23	±14
All observers	±20	±17

The first column refers to the probable errors of observations already corrected for (slowly changing) personal equations. The probable error of the original observations, including the effect of unknown personal equations, is ±32 msecs. It should be stressed that the figures in the second column represent upper limits; they are, in fact, the values of $0{\cdot}6745\sigma_{\epsilon'}$, where $\epsilon' = e + (2C - l)\sec\phi$. They thus represent upper limits to the value of $0{\cdot}6745\sigma_e$ and incorporate the probable error of the real variations C of collimation about the steady drift, which we have assumed zero in deducing the figures. And, of course, they also include contributions due to errors of star places and clock erratics, the effects of these having been tacitly included in our σ_e.

8. The level errors have now to be considered. A straight determination of the correlation between clock and level errors is not the method used in practice. The level errors involve the collimation readings because they are obtained by differencing the nadir readings of the R.A. micrometer against the measured reading for the line of collimation. Consequently we have used the level errors with the collimations subtracted, that is the nadir readings. This subtraction is justified only if the removed collimations have approximated closely to a linear drift. For instance, if the collimation suffered an abrupt change the subtraction would introduce a discontinuity into the nadir readings. The figures, in terms of revolutions of the R.A. micrometer, were converted into seconds of time at the Edinburgh zenith by multiplying by $1{\cdot}306\sec\phi = 2{\cdot}33$.

The nadir readings are subject to a variation with temperature. Accordingly, it was necessary to determine this temperature coefficient, in order to derive the nadir readings of level which would have been obtained if the temperature had remained steady.

Unfortunately, it has not been the practice to record the temperatures at the time of observation. The only records available at the Observatory were the routine meteorological readings taken at 21^h U.T. These have been supplemented by thermograph records kept by Mr J. Paton at the Physics Laboratory of the University of Edinburgh, and Mr Paton has kindly allowed Miss Angela Penny (sister of one of the authors) to extract the necessary information from the thermograph sheets. These figures have been used to correct the Observatory readings at 21^h U.T. to the mean times of observation of the transits. The temperatures so obtained are relatively unreliable and only a rough approximation to the temperature coefficient can be hoped for. Least square solutions were

made for straight lines defining the temperatures and nadir readings separately when plotted against the date, and the residuals from each of these straight lines were then plotted against each other on a scatter diagram. The correlation of the residuals was not well defined but was considered sufficiently definite to give us confidence that the nadir readings would be smoothed somewhat by correction for temperature. An added difficulty was that the distribution of observers with temperature was not uniform, and as there was some indication of systematic difference in the levels as measured by the different observers, it was not enough merely to determine the slope of a line drawn through the points on the scatter diagram irrespective of observer. It was necessary to adopt different lines for the different observers but with a common slope. The adoption was made by a least square solution, and the temperature coefficient obtained was $+2{\cdot}8$ milliseconds, at the zenith, per degree Fahrenheit.

The coefficient was determined many years ago by Professor Sampson * who obtained $+8{\cdot}5$ equatorial milliseconds per degree Fahrenheit for the level error temperature coefficient. To compare this result with ours we must multiply it by $\sec\phi = 1{\cdot}785$ so as to convert to milliseconds at the Edinburgh zenith. This gives $+15{\cdot}2$, as against our $+2{\cdot}8$. It should be observed, however, that Sampson's result is for temperatures recorded by a thermometer with the bulb sunk some inches into one of the transit circle piers, whereas the temperatures we are using are referred to readings in a Stevenson screen. It is to be expected that Sampson's result would be more definite and would give a larger coefficient since temperature fluctuations of comparatively short period would be minimized on the pier as against the variations in the screen. Since we have only screen temperatures available we must naturally use a temperature coefficient applicable to them. The nadir readings for level (i. e. the level errors with the collimations subtracted and the resulting differences multiplied by $\sec\phi$) were accordingly corrected for the temperature coefficient $+2{\cdot}8$ and some smoothing resulted.

At this stage the corrected nadir readings were reversed in sign so that a positive correlation with clock errors was to be expected as a result of the common errors of measurement of the nadir settings. As in the case of collimations, it was necessary to adopt separate datum lines for the different observers to allow for any secular change in the personal equations in the measurement of level error. The lines were defined by least square fits of expressions $a+bt$ to the numbers, one such fit being made for each observer. Residuals from the datum lines were then deduced.

These nadir residuals were not analysed for correlation with the original clock error residuals but with the clock error residuals from whch the collimation residuals had been subtracted, i. e. with the clock error residuals obtained on the assumption that the collimation drifted linearly with time.

We shall, as before, investigate the significance of this correlation coefficient before giving the numerical results.

To obtain the residuals of clock error corrected for collimation, subtract the quantities $-2\chi\sec\phi$ from the quantities ϵ, thereby forming the quantities

$$\epsilon' = \epsilon + 2\chi\sec\phi = e + (2C - l)\sec\phi.$$

To obtain the nadir residuals, subtract the quantities χ from the quantities λ,

* *M.N.*, **75**, 69, 1914.

forming the quantities

$$\lambda' = \lambda - \chi = L + l - C.$$

The coefficient r_2 is defined as the coefficient of correlation between ϵ' and $-\lambda' \sec\phi$, so that

$$r_2 = -\frac{\sigma_{\epsilon'\lambda'}}{\sigma_{\epsilon'}\sigma_{\lambda'}},$$

where
$$\sigma_{\epsilon'\lambda'} = -(2\sigma_C^2 + \sigma_l^2)\sec\phi, \tag{4}$$

assuming that the other sums of products vanish as before,

also
$$\sigma_{\epsilon'}^2 = \sigma_e^2 + (4\sigma_C^2 + \sigma_l^2)\sec^2\phi, \tag{5}$$

and
$$\sigma_{\lambda'}^2 = \sigma_C^2 + \sigma_L^2 + \sigma_l^2. \tag{6}$$

The numerical values of the coefficient of correlation between the nadir and corrected clock error residuals are:

Observer	r_2 and p.e. of r_2
LS	$0{\cdot}07 \pm 0{\cdot}19$
B	$0{\cdot}78 \pm 0{\cdot}19$
JP	$0{\cdot}14 \pm 0{\cdot}17$
MM	$-0{\cdot}13 \pm 0{\cdot}21$
All observers	$0{\cdot}10 \pm 0{\cdot}10$

For three observers these figures do not provide definite evidence of correlation. For the observer B the correlation coefficient 0·78 suggests strong correlation, but there are only six observations by this observer over the period considered. However, the probable error of $\pm{\cdot}19$ suggests that the coefficient obtained cannot be entirely due to sampling error and an examination of the six pairs of residuals involved strengthens our belief that there is a pronounced correlation for B. These residuals are as follows:—

Residuals, in units of one millisecond

Of clock error after correction for collimation.	−17	+ 6	− 4	+34	−3	−16
Of the nadir readings.	−10	+20	−16	+18	−3	− 9

Quite apart from the actual residuals and correlation coefficients there is reason to believe that the correlation should be stronger in the case of observer B. For any observer there must necessarily be some correlation since the quantity σ_l cannot be zero but the correlation will be small if σ_l and σ_C are much smaller than one of the other standard deviations involved in (5) and (6). Now one of these is σ_L, the standard deviation of the real variations of the level error about the datum taken as the "mean value". But this datum involves the temperature, and we have already referred to the unsatisfactory nature of the temperatures used. It is, we think, significant that the observations of the three observers LS, JP and MM were made over a temperature range twice as large as that for the observations of B. The suggestion is that for the three observers σ_L is enhanced and the enhanced value reduces the correlation coefficient whereas for B σ_L is smaller. This suggestion receives support from the table of collected results given below in Section 11; from this table it appears that the probable error of the nadir measures, corrected for temperature, is, in fact, nearly twice as large for the three observers as it is for B.

It should be mentioned that a possible additional complication would be a variation of *collimation* with temperature. This would introduce a correlation

between C and L and such a correlation would give rise to additional terms in r_2 which would be enhanced for the three observers LS, JP and MM as compared with observer B on account of the smaller temperature range for B's observations. But without knowledge of the sign of the possible correlation between C and L it is impossible to say whether its effect would be to decrease or increase the correlation coefficients.

Since the correlation coefficients r_2 for the three observers LS, JP and MM are small, we cannot obtain reduced upper limits to that part of the probable error of a time determination not due to erroneous measures of instrumental errors in a way similar to our treatment of collimation. For observer B we can obtain such a reduced upper limit by differencing the two columns of residuals quoted above, thereby obtaining the quantities $\epsilon'' = \epsilon' + \lambda' \sec\phi = e + (L + C)\sec\phi$. For this observer the probable error of a time determination corrected for erroneous collimation measures is $0{\cdot}6745\ \sigma_{\epsilon'} = \pm 14$ msecs. (see Section 7); it is reduced, when corrected for error of measurement of the nadir settings to $0{\cdot}6745\ \sigma_{\epsilon''} = \pm 9$ msecs. These figures are, from what has been said, upper limits to $0{\cdot}6745\ \sigma_e$, but it must be remembered that they are subject to sampling error which is necessarily enhanced for a sample of only six observations.

9. By forming the products of the collimation and nadir residuals we obtain a further mean product,

$$\sigma_{\chi\lambda'} = -\sigma_C^2. \qquad (7)$$

This supplies a value of σ_C, the standard deviation of the real excess of the collimation over the adopted linear drift, assuming, as before, that the mean products of the components vanish.

This completes a set of seven equations involving all the unknown quantities we have been discussing.

In the following table are collected the numerical values corresponding to equations (1) to (7) for the mean of the four observers:—

		(msecs.)	
M.P. of C.E. × Colln.			
$= \sqrt{-2\sigma_{\epsilon\chi}\sec\phi}$	$= 2\sigma_c \sec\phi$	$\times 0{\cdot}6745 = \pm 14$	(i)
S.D. of C.E.			
$= \sigma_\epsilon$	$= \sqrt{\sigma_e^2 + (4\sigma_c^2 + \sigma_l^2)\sec^2\phi}$	$= \pm 20$	(ii)
S.D. of Collimation			
$= 2\sigma_\chi \sec\phi$	$= 2\sqrt{\sigma_C^2 + \sigma_c^2}\sec\phi$	$= \pm 16$	(iii)
M.P. of (C.E. − Colln.) × Nadir			
$= \sqrt{-\sigma_{\epsilon'\lambda'}\sec\phi}$	$= \sqrt{2\sigma_C^2 + \sigma_l^2}\sec\phi$	$= \pm 6$	(iv)
S.D. of (C.E. − Collimation)			
$= \sigma_{\epsilon'}$	$= \sqrt{\sigma_e^2 + (4\sigma_C^2 + \sigma_l^2)\sec^2\phi}$	$= \pm 17$	(v)
S.D. of Nadir			
$= \sigma_{\lambda'}\sec\phi$	$= \sqrt{\sigma_C^2 + \sigma_L^2 + \sigma_l^2}\sec\phi$	$= \pm 21$	(vi)
M.P. of Colln. × Nadir			
$= \sqrt{-2\sigma_{\chi\lambda'}\sec^2\phi}$	$= \sqrt{2}\,.\,\sigma_C \sec\phi$	$= \pm 11$	(vii)

where C.E. stands for clock error. The values are multiplied by 0·6745 to put them on the same footing as probable errors, used throughout the rest of the paper.

Theoretically a solution of best fit could be obtained giving all the probable errors involved, and this would be feasible provided the number of observations was big enough. But for an effective number of 43 observations the sampling errors are not small enough to justify any attempt at a rigorous solution. Apart from this the table indicates the approximate magnitude of the various quantities, some of which have already been quoted.

Four of the equations can be used to obtain a rough measure of the quantity $0{\cdot}6745\ \sigma_e$.

From (i) and (ii) we get $0{\cdot}6745\sqrt{\sigma_e^2+\sigma_l^2\sec^2\phi}=\pm 15$.

From (v) and (vii) we get $0{\cdot}6745\sqrt{\sigma_e^2+\sigma_l^2\sec^2\phi}=\pm 7$.

The disagreement illustrates the effects of sampling error and of the assumption that there is no correlation between the various components of the collimation and level measures. It can be taken that $0{\cdot}6745\sqrt{\sigma_e^2+\sigma_l^2\sec^2\phi}=\pm 15$ is an upper limit. Therefore, $0{\cdot}6745\ \sigma_e$, the probable error of that part of the scatter in clock error not due to the effect of erroneous measures of collimation and level, is less than ± 15 msecs. It would require much more extended data to arrive at a more definite conclusion by these methods.

10. Approaching the problem from a totally independent direction, values for $0{\cdot}6745\sigma_e$ can be obtained from the internal agreement of the results from individual stars for the same night's work.

On a given night 10 stars are usually employed and the probable error of the resulting clock error can be deduced from the residuals of the individual stars from the line of best fit in the plot against $\tan\delta$, of $\alpha - T$ (corrected for collimation and level), where α denotes R.A. and T the observed time of transit of a star. The probable errors obtained in this way are practically independent of the effects of personal equation and of erroneous measures of the instrumental errors. The values for each observer, and for the mean of all four observers are :

LS	$\pm$ 7 msecs.
B	$\pm$ 9
JP	$\pm$ 8
MM	$\pm$10
All Observers	$\pm$ 9

For three observers these values are similar to, though rather smaller than, the values obtained by correcting the clock errors for erroneous measures of collimation and level. The values of $0{\cdot}6745\ \sigma_e$, derived in this way, may be expected to be lower limits to the true values, because errors, such as would be caused for example by a lateral refraction, which are common to the observations during a night, are eliminated. If, therefore, the larger values, derived in Section 7, are accepted as most nearly representing the true values of the probable errors, they are seen to be considerably smaller than the total probable error of a measure of clock error (± 32 msecs.).

11. The results for the individual observers so far obtained are given in the following table (referred to in Sections 7 and 8):—

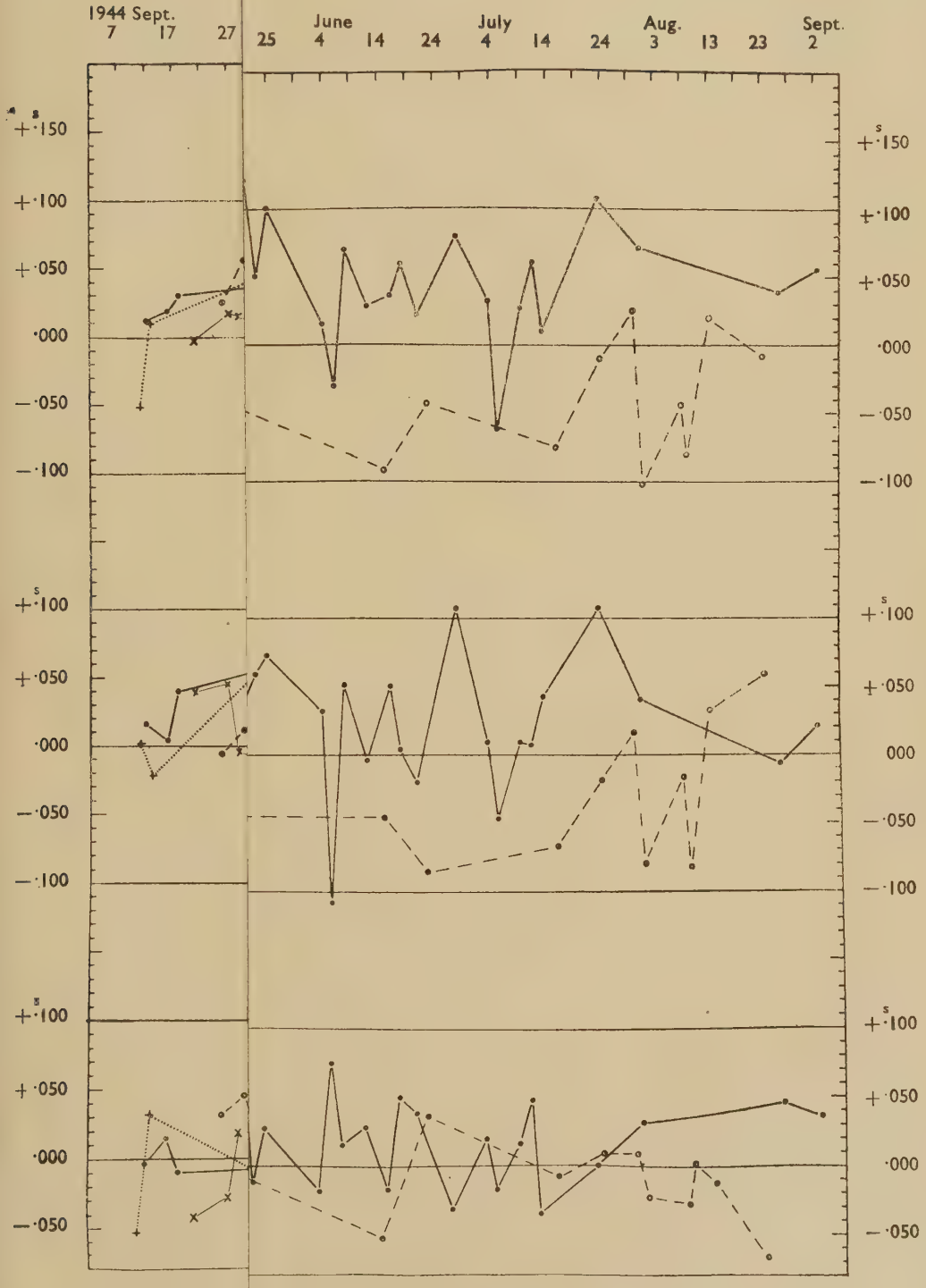
Observer	Collimation	Nadir	Probable Errors (msecs.) Clock Error: Uncorrected (All Observers)	Corrected for personal equations	Corrected for collimation	Corrected for level	From internal agreement
LS	±15	±22	±32	±16	± 9		± 7
B	±11	±12		±18	±14	± 9	± 9
JP	±15	±22		±22	±23		± 8
MM	±19	±20		±23	±14		±10

12. We have been forced to confine the analysis to the short period October 19 to February 20 for which the collimation error did not seem to be subject to appreciable discontinuities, and we have seen that the effect of erroneous measures of collimation betrays itself by a strong correlation between collimation and clock errors. Although detailed analysis is not feasible for the remainder of the whole period covered by the data, it is easy to show by a graphical method that the correlation persists.

The heavy continuous line in Fig. 1 (Plate 11) over the period October 19 to February 20 represents a datum line for the mean of all four observers. The heavy continuous line extending from February 1 to September 3 (over the greater part of this period only two observers, LS and JP, shared in the transit work) represents a datum line for the mean of the two observers LS and JP. The thin lines represent least square solutions for these two observers for the period February 1 to May 5 and May 6 to September 3, and the heavy continuous line is the mean of these, but adjusted so that the two portions intersect at May 5, one of the dates on which a change of rate of IVC took place.

The residuals of the clock errors for the different observers from the heavy continuous lines for the period October 19 to Feburary 20 and February 20 to September 3, are plotted in Fig. 3 (Plate 12), and this figure also exhibits a plot of the collimation residuals reckoned from the datum lines for collimation shown in Fig. 1 (Plate 11). On comparing the two graphs it is at once apparent that the fluctuations in the collimation graph are reproduced in the graph of clock error residuals, that is that there is a strong correlation between clock error and collimation over the whole period.

The graphs also appear to show that there is a correspondence between the general trends of collimation and clock errors over the rest of the whole period for the two observers LS and JP, that is that the varying difference of personal equation in collimation is reflected in the varying difference of personal equation in clock error. But it must be pointed out that we have really preserved the feature by the method used in the adoptions of the collimation datum lines shown in Fig. 1 (Plate 11). This procedure will, however, not produce the detailed day-to-day correlation which is a striking feature of the graphs and which indicates that the erroneous measures of collimation contribute materially to the total scatter in clock error.



L. S. T. Symms and C. J. A. Penny, Errors of Time Determination.

At the bottom of Fig. 3 (Plate 12) there appears a plot of the clock error residuals with the collimation residuals subtracted. In the case of the observer B the level residuals have also been subtracted. It will be seen that the ranges of the fluctuations shown in the clock residual graph are materially decreased.

13. It has long been suspected that the major part of the probable error of a determination of time with a transit circle of classical pattern is due to the effect of erroneously determined instrumental constants and that the contribution made by the actual errors of observing the transits of the stars is relatively small. All the numerical evidence in this paper goes to confirm this suspicion. It is very well brought out in the case of the two observers with long observing experience:

Probable errors of clock errors (msecs.)

Observer	Uncorrected (from whole series)	Corrected for personal equations	Corrected for collimation	Corrected for level	From internal agreement
LS	±32	±16	± 9	...	±7
B		±18	±14	±9	±9

It should be noted that these probable errors are applicable to a hand-driven impersonal micrometer and include the effects of errors in the individual star places.

Royal Observatory,
Greenwich:
1946 *May* 23.

Royal Observatory,
Edinburgh.

ON THE CONDENSATION OF THE PLANETS

F. Hoyle

(Received 1946 April 13)

Summary

The main object of the present paper is to consider the condensation of planets in a gaseous disk of material moving in orbital motion about the Sun. The following results are obtained :—

(1) Condensations grow in much the same manner as droplets form in a supersaturated atmosphere of water vapour. This process is mainly important for condensations with masses less than about 10^{23} gm. This mass is attained, by condensations that do not suffer many collisions with other condensations, in a time of order 10^{8} years.

(2) For condensations with masses greater than 10^{23} gm. the growth is mainly by accretion. The time taken for the mass to increase from 10^{23} gm. to a value of order 10^{30} gm. is also about 10^{8} years. The final stages of the accretion process are very rapid. Thus a condensation requires only about 10^{5} years for growth from 10^{28} gm. to 10^{30} gm.

(3) The rapidity of the final stages of growth by accretion means that the first condensations to acquire masses of order 10^{28} gm. will quickly go on to attain masses comparable with the total mass of the planetary material. This result has two important consequences.

(i) That a large fraction of the gaseous material will condense into a few bodies of large mass rather than into a swarm of small particles. Thus only a small fraction of the gaseous material remains finally in the form of condensations of small mass. These small condensations are associated with the asteroids.

(ii) The terrestrial planets and satellites cannot be formed by condensation. The reason for this is that the mass of a condensation does not stop at a value of order 5×10^{27} gm. but must quickly go on to values of the order of the masses of the major planets. This result confirms Lyttleton's theory of the formation of the terrestrial planets and satellites.

(4) The axial rotation period of planets of large mass is shown to be of order 5 hours. This result provides further confirmation of Lyttleton's theory, since in this theory it is necessary for the primitive planets to possess rotation periods less than about 7 hours.

1.1. *Preliminary Remarks.*—The problem of the origin of the planets falls naturally into two parts:—

(i) To show how the Sun came to acquire the planetary material.

(ii) Since the planetary material must almost certainly have been gaseous in the first place, it is necessary to discuss the condensation of the material into solid bodies.

It is important to notice in connection with (i) that it is most unlikely that any form of accretion of interstellar material could have been responsible for the origin of the planetary material. For there seems to be no such process that could produce the necessary concentration of material. Thus the density of interstellar material is not likely to exceed 10^{-19} gm. per cm.^{3}, whereas the required density of the planetary material is of order 10^{-12} gm. per cm.^{3}. The latter density is obtained by distributing the total mass of the planets uniformly through a disk with a ratio

of thickness to radius of 1/10, the radius of the disk being taken as equal to the diameter of Jupiter's orbit.

The above remarks indicate that we must seek a theory in which the solar system is formed from material detached from a star. The first outstanding point that arises in searching for such a theory is that angular momentum considerations make it clear that the star supplying the planetary material cannot be the Sun.* The writer has recently proposed a theory in which the Sun is supposed to have once possessed a binary companion that became a supernova.† This suggestion is similar in principle but differs in detail from a theory proposed previously by Lyttleton.‡ The three essential features of the supernova theory are:—

(1) It requires only a very slight deviation from strict symmetry in the way material is thrown off by the supernova during the outburst to produce a recoil on the remaining stellar nucleus sufficient to remove it from the solar system.

(2) The Sun is able to acquire permanently a small fraction of the material thrown off during the outburst.

(3) The average angular momentum per unit mass of the captured material about the Sun is to be equal to the average angular momentum per unit mass possessed by the planets.

The main advantages of this process are firstly that it is known by observation to occur. Secondly, in any catastrophic theory the general velocities of ejection of diffuse material are so high that the capture of material by the Sun is almost certain to be an inefficient process. That is, it seems likely that the Sun can only acquire a small fraction of the total amount of diffuse material removed from the star in question. This means that a quantity of material with a mass of *stellar* rather than planetary order must be thrown off from the star. The removal of such a quantity of material involves an outburst of an intensity that is only encountered in supernovae.§ It follows from these remarks that the supernova theory enables a satisfactory discussion to be given of part (i) of the theory of the origin of the planets. It remains to investigate the condensation of diffuse material into solid bodies.

1.2. The diffuse material acquired by process (i) forms a gaseous envelope around the Sun. On account of its angular momentum the material moves in orbits that are mainly situated at distances between 5 and 10 astronomical units from the Sun. The mean free path for random motions of the particles in the gas is about a kilometre. Since this value is very small compared with the dimensions of the orbits it can be shown that any initial eccentricities or variations in the inclinations of the orbits of the particles in the material must quickly be damped

* H. N. Russell, *The Solar System and its Origin*, New York, 1935 ; R. A. Lyttleton, *M.N.*, **98**, 536, 1938.

† F. Hoyle, *Proc. Camb. phil. Soc.*, **40**, 256, 1944 and *M.N.*, **105**, 175, 1945.

‡ R. A. Lyttleton, *M.N.*, **101**, 216, 1941.

§ This remark is apparently similar to Luyten's criticism of Lyttleton's original collision theory. There is, however, an important difference, for until Spitzer (*Ap. J.*, **90**, 675, 1939) had shown that the diffuse planetary material cannot condense into compact planetary bodies *before* capture by the Sun takes place it was possible to defend the collision theory by supposing that the condensation of the planets takes place almost immediately after the material leaves the parent star. If such a condensation could take place there is a chance of about one in a hundred that the Sun captures practically the *whole* of the condensed material. This situation is markedly different from the case where the material remains gaseous, for in this case the Sun can capture only about one per cent. of it.

out. After a few revolutions the material comes to move everywhere in approximately circular orbits about an axis through the Sun in the direction of the total angular momentum vector of the material. When this stage is reached the temperature of the material becomes close to $T_s(R_\odot/r)^{\frac{1}{2}}$, where T_s, $R_\odot$ are the surface temperature and radius of the Sun, and r is the mean distance of the material from the Sun. (At a density of 10^{-12} gm. per cm.^{3} the material is markedly opaque to radiation and the frequency of the radiation emitted by the Sun is degenerated by repeated absorption and re-emission. A steady state must eventually be reached in which the rate of radiation of the planetary material into outer space is equal to the solar luminosity $L_\odot$. This requires the temperature T of the material to have an average value such that $aT^4 \simeq L_\odot/\pi r^2$.) Thus for r equal to the radius of Jupiter's orbit the temperature is of order 200° K. As it is preferable from the point of view of Section 2 that the temperature be overestimated rather than underestimated, a standard value of 273° K. will be used in the work given below. Since the thermal velocities of all atoms at this temperature are small compared with $(GM_\odot/r)^{\frac{1}{2}}$ it follows that the planetary material must form a thin disk. If z is the distance of a point from the central plane of the disk the gravitational acceleration due to the Sun in a direction perpendicular to the plane of the disk is given to sufficient accuracy by $GM_\odot z/r^3$, where r is the distance of the point from the Sun. Thus for material at uniform temperature T the equation of hydrostatic equilibrium in a direction perpendicular to the plane of the disk is

$$\frac{\mathfrak{R}T}{\mu}\frac{\partial \rho}{\partial z} = -\frac{GM_\odot \rho z}{r^3},$$

where μ is the mean molecular weight and ρ is the density of the gaseous material, $M_\odot$ is the solar mass, G is the constant of gravitation, and $\mathfrak{R}$ is the gas constant. This equation integrates immediately and shows that

$$\frac{\tau}{r} \simeq \left(\frac{2\mathfrak{R}Tr}{GM_\odot \mu}\right)^{\frac{1}{2}}, \tag{1}$$

where τ is an estimate for the thickness of the disk. Equation (1) gives $\tau/r = 0{\cdot}056$ when $T = 273°$ K., $r = 10^{14}$ cm., and $\mu = 16$. The plane of the disk is perpendicular to the total angular momentum vector of the planetary material. A further point that may be noted is that gas pressure $\mathfrak{R}\rho T/\pi$ satisfies the inequality

$$\frac{\mathfrak{R}\rho T}{\mu} \ll \frac{GM_\odot \rho}{r}, \tag{2}$$

where ρ is the density of the gaseous material and $\mathfrak{R}$ is the gas constant.

1.3. It is well known that gravitational instability cannot produce condensations in the rotating disk of gaseous material. The reason for this is easily understood. Consider the diffuse material in a spherical region of radius s lying wholly within the disk ($s < \tau$). The gravitational acceleration $4\pi G\rho s/3$ arising from the material in this region tends to cause contraction but this tendency is opposed by the gravitational field of the Sun which produces a differential acceleration throughout the condensation of order $GM_\odot s/r^3$. Condensation can occur only if the internal gravitational acceleration $4\pi G\rho s/3$ exceeds the differential acceleration of the Sun. This is impossible in the present problem, for it would require

$$4\pi r^3 \rho/3 > M_\odot, \tag{3}$$

which certainly cannot be satisfied since $4\pi r^3\rho/3$ is only of the order of the mass of the planets.* It follows therefore that gravitational condensation is not possible owing to the disruptive effect of the Sun's field.

An important step towards obtaining a satisfactory theory of the condensation of the planets has recently been made by Parson †, who, following a suggestion of Jeffreys ‡, has shown that the partial pressures of such gases as iron, silicon dioxide, magnesium oxide, etc. actually existing in the diffuse gaseous material, must exceed by a very large factor the equilibrium vapour pressures for a temperature of 273° K. of solids composed of these substances. Thus solid bodies must tend to form in the diffuse planetary material in a manner analogous to the formation of raindrops in a supersaturated atmosphere. In view of the significance of this result it seems desirable to give an independent estimate for the equilibrium vapour pressure of a typical substance such as iron. This question is considered in Section 2 where the methods of statistical mechanics (in place of the thermodynamic treatment of Parson) are shown to lead to a result in substantial agreement with that obtained by Parson.

The result that solid bodies must form in the diffuse planetary material by direct condensation does not in itself provide a satisfactory solution of the problem of the condensation of the planets. Two further questions are considered in the present paper:—

(1) In addition to showing that solid bodies form by direct condensation it is also necessary to show that the condensations can grow to the required dimensions in a time of the order of, or less than, 10^9 years.

(2) In Jeans's tidal theory and also in the Jeffreys and Lyttleton collision theories it has been assumed that the planetary material condenses almost immediately after leaving the parent star. One of the objects of this assumption was to enable the observed axial rotations of the planets to be obtained. A new question arises, now that this assumption is abandoned, as to whether planets condensing in the disk of planetary material will possess the required axial rotations.

These two questions are considered in Sections 3 and 4 respectively.

2. *The Equilibrium Vapour Pressure of Iron.*—It is sufficiently accurate for the present purpose to neglect the internal structure of the iron. That is, we regard iron as being composed of structureless massive points in both the gaseous and crystalline phases. The errors introduced by this assumption are completely negligible compared with other factors that arise in the present problem (the errors introduced cannot be greater than a factor of about 10 at most, as compared with the factor of order 10^{25} discussed below). With this assumption the equilibrium vapour pressure is given by §

$$\frac{(2\pi m_{\mathrm{Fe}} kT)^{\frac{3}{2}}}{h^3} \cdot kT \exp\left[-\frac{X}{kT} - \int_0^T \frac{dT'}{\mathfrak{K}T'^2}\int_0^{T'} (C_p)_{\mathrm{sol}}\, dT''\right], \qquad (4)$$

* This difficulty does not occur for the condensation of a *star* in a nebula. It is easy to see that in this case the condition corresponding to (3) is

$$\text{Density of interstellar material} > \frac{\text{Gravitational acceleration due to nebula}}{\text{Radius of nebula.}}.$$

For a typical nebula the gravitational acceleration is of order 10^{-7} c.g.s. units, so that even for a nebula with a radius as small as 10^{21} cm. this inequality is satisfied by a large margin.

† A. L. Parson, *M.N.*, **105**, 244, 1945.

‡ H. Jeffreys, *Nature*, **153**, 140, 1944.

§ R. H. Fowler, *Statistical Mechanics*, p. 173, Cambridge, 1936.

where m_{Fe} is the mass of an atom of iron, h is Planck's constant and $(C_p)_{sol}$ is the specific heat at constant pressure of crystalline iron. The quantity NX is the energy required to evaporate completely a crystal containing N iron atoms at the absolute zero of temperature. The crystal is regarded as being sufficiently large for surface effects to be neglected. That is, the quantity X is independent of N. Thus if we define H by

$$H = \mathfrak{R}X/k \tag{5}$$

then H is the heat of evaporation per mole at the absolute zero of temperature.

The quantities appearing in the factor multiplying the exponential in (4) are all known constants. By using the values of these constants when $T = 273°$ K. this factor has a numerical value of $1{\cdot}73 \times 10^{13}$. Thus at $T = 273°$ K. the equilibrium vapour pressure is

$$1{\cdot}73 \times 10^{13} \exp\left\{ -\frac{H}{\mathfrak{R}T} - \frac{1}{\mathfrak{R}} \int_0^T \frac{dT'}{T'^2} \int_0^{T'} (C_p)_{sol}\, dT'' \right\} \text{dynes per cm.}^2. \tag{6}$$

To effect further progress it is necessary to know the value of H. Unfortunately no experimental determination seems to be available and an assumption has to be made. For comparison with Parson's work the assumption of Trouton's rule will be used. Trouton's rule states that the heat of evaporation *at boiling-point* of a substance is $21 T_b$ cal. per mole, where T_b is the boiling-point in °K. For iron $T_b = 2723°$ K. and this gives a heat of evaporation at boiling-point of 57183 cal. per mole. Now since the thermal energy per unit mass is greater at boiling-point than at the absolute zero of temperature, it follows immediately that the heat of evaporation at absolute zero must be greater than the heat of evaporation at boiling-point. Thus on the basis of Trouton's rule it follows that

$$H > 57183 \text{ cal. per mole.}$$

Accordingly we have

$$\exp\left\{ -\frac{H}{\mathfrak{R}T} - \frac{1}{\mathfrak{R}} \int_0^T \frac{dT'}{T'^2} \int_0^{T'} (C_p)_{sol}\, dT'' \right\} < \exp\left(-\frac{H}{\mathfrak{R}T} \right) < \exp\left(-\frac{57183}{\mathfrak{R}T} \right) \simeq 10^{-46} \tag{7}$$

when $T = 273°$ K. From inequality (7) applied to (6) we obtain an *upper limit* of $1{\cdot}73 \times 10^{-33}$ dyne per cm.^{2} $= 1{\cdot}73 \times 10^{-39}$ atmospheres for the vapour pressure of iron at $T = 273°$ K. This vapour pressure is extremely small compared with the partial pressure of iron existing in the diffuse planetary material. Thus if iron forms only one per cent. by mass in material of density 10^{-12} gm. per cm.^{3} the partial pressure is 4×10^{-12} atmospheres.

The above value of $1{\cdot}73 \times 10^{-39}$ atmospheres is larger than the corresponding estimate of 2×10^{-47} atmospheres obtained by Parson. This difference is due chiefly to the neglect in the present work of the difference between the heats of evaporation of iron at boiling-point and at the absolute zero of temperature. The neglect of the integral

$$\int_0^T \frac{dT'}{T'^2} \int_0^{T'} (C_p)_{sol}\, dT''$$

is less important. For $T = 273°$ abs. the specific heat is about 6 cal. per mole per degree, and the integral has a magnitude of order 6 which is very small compared with H/T. (It may be noted that for small T'' the specific heat $(C_p)_{sol}$ is proportional to $(T'')^3$. Thus the integral does not diverge at the lower limit. This statement applies to the specific heat of the metal lattice. There is also a very small term proportional to T'' arising from the contribution of the electrons in the metal.)

The only important assumption in the above work is the use of Trouton's rule for estimating H. The factor by which the partial pressure of iron exceeds the vapour pressure calculated above is so great, however, that a very large deviation from Trouton's rule can be tolerated without upsetting the validity of Parson's main conclusion. For example, even if the heat of evaporation at boiling-point is as little as $10T_b$ cal. per mole (as opposed to $21T_b$ cal. per mole given by Trouton's rule) the calculated vapour pressure is still less than 10^{-12} atmospheres.

The conclusions reached above evidently apply to all planetary constituents with boiling-points close to, or greater than, the boiling-point of iron.

3. *The Rate of Growth of the Planetary Condensations.*

3.1. *The Direct Condensation Process.*—The next step in the present investigation is to show that the process discussed above is capable of producing condensations of planetary dimensions in a time of the order of, or less than, 10^9 years. To begin with, the rate of growth of a condensation will be calculated on the assumption that the condensation is unaffected by collisions with other condensations. Since the partial pressures of the main planetary constituents exceed the saturation vapour pressures of solid bodies composed of these substances, it follows that the rate of increase in mass of a condensation that does not suffer collisions with other condensations is given in terms of the rate at which gaseous material strikes and adheres to the surface of the condensation. The condensation acts as a hydrodynamic sink and the rate at which material strikes the surface is

$$4\pi s^2\rho \times [\text{Mean thermal velocity of the gaseous atoms}],$$

where s is the radius of the condensation. Accordingly we obtain the equation

$$\frac{d\mathscr{M}}{dt} \simeq 4\pi s^2 \rho f \left(\frac{3kT}{\mu m_H}\right)^{\frac{1}{2}}, \tag{8}$$

where $\mathscr{M}$ is the mass of the condensation and f is the probability of adherence. The question of the numerical value of f is discussed by Fowler * who concludes that f is likely to be close to unity. The present theory is not sensitive to the value chosen for f.

The radius s increases to $(s+\delta s)$ in a time δt given by

$$4\pi s^2 \delta_s \sigma = \frac{d\mathscr{M}}{dt} . dt, \tag{9}$$

where σ is the mean density in the condensation.

Substituting for $d\mathscr{M}/dt$ from (8) gives

$$\frac{ds}{dt} = \frac{\rho f}{\sigma}\left(\frac{3kT}{\mu m_H}\right)^{\frac{1}{2}} \tag{10}$$

which integrates immediately to give

$$s - s_0 = \frac{\rho f}{\sigma}\left(\frac{3kT}{\mu m_H}\right)^{\frac{1}{2}} t, \tag{11}$$

where s_0 is the radius of the condensation at time $t=0$. The time taken for the condensation to grow from a small radius ($s_0 \ll s$) is given by

$$\frac{\sigma s}{\rho f}\left(\frac{\mu m_H}{3kT}\right)^{\frac{1}{2}}. \tag{12}$$

At first sight it would seem that an estimate for the time of condensation of the planets can be obtained by inserting the present planetary radii in (12). This is

* *Loc cit.*, p. 698.

not correct, however, because a new process intervenes when s rises to a critical value. Thus the present process of direct condensation applies only so long as the velocity of escape $\sqrt{(2G\mathscr{M}/s)}$ from the surface of the condensation is less than the mean thermal energy $(3kT/\mu m_H)^{\frac{1}{2}}$ of the gaseous atoms. When the velocity of escape exceeds $(3kT/\mu m_H)^{\frac{1}{2}}$ equation (8) no longer applies and must be replaced by the rate of accretion of a compact body in a gas. It is seen therefore that the value of s given by

$$2G\mathscr{M}/s = 3kT/\mu m_H \tag{13}$$

represents a critical value that separates growth by direct condensation from growth by accretion. In the present subsection we are only concerned with growth by direct condensation. The question of accretion will be considered in 3.2.

If we put $\mathscr{M} = 4\pi s^3\sigma/3$ in (13) we obtain

$$s = \left(\frac{9kT}{8\pi G\mu\sigma m_H}\right)^{\frac{1}{2}}. \tag{14}$$

The time taken for growth by direct condensation is given immediately on substituting this value of s in (12). This gives

$$\frac{\sigma}{\rho f}\left(\frac{3}{8\pi G\sigma}\right)^{\frac{1}{2}} = 8{\cdot}9 \times 10^7 \text{ years} \tag{15}$$

when $\sigma = 4$ gm. per cm.^{3}, $\rho = 10^{-12}$ gm. per cm.^{3} and $f = 1$.

3.2. *Growth by Accretion.*—The calculations of the present subsection will be made on the assumption that the condensation moves in a circular orbit about the Sun. The accretion process in this case differs in an important respect from previous discussions *, for in former work the bulk motion of material relative to the accreting body was always large compared to the thermal velocities of the atoms in the material, whereas in the present problem the thermal velocities are large compared with the bulk velocites. Nevertheless it can be seen that accretion takes place in much the same way in the two cases. This can be seen by considering the following simple model. Let material at uniform temperature T be moving radially inwards towards the condensation and let the inward velocity at distance x from the centre of the condensation be $(2G\mathscr{M}/x)^{\frac{1}{2}}$. Then the equation of continuity shows that the density of the material depends on x according to the factor $x^{-\frac{3}{2}}$. Thus the outward pressure gradient in $3\mathfrak{R}\rho T/2\mu x$ which may be compared with the gravitational force due to the condensation $G\rho\mathscr{M}/x^2$. These two forces are equal in magnitude and opposite in sign when

$$x = \frac{2G\mathscr{M}\mu}{3\mathfrak{R}T}. \tag{16}$$

When x is less than the right-hand side of (16) the inward gravitational force is greater than the outward pressure gradient and our model is approximately valid, but the model breaks down when x exceeds (16) since the gravitational field of the condensation is then unable to control the motion of the material. This means that we have a hydrodynamical sink with a radius equal to (16) instead of equal to the radius of the condensation as in 3.1. (This statement is subject to the proviso that the differential effect of the Sun's field must be negligible. This question is discussed in detail below.) Thus the rate at which material is added to the

* F. Hoyle and R. A. Lyttleton, *M.N.*, **101**, 227, 1941 ; H. Bondi and F. Hoyle, *M.N.*, **104**, 273, 1944.

condensation is given by omitting the factor f in (8) and by replacing the radius s of the condensation by the right-hand side of (16). This gives

$$\frac{d\mathscr{M}}{dt} \simeq 4\pi G^2 \mathscr{M}^2 \rho \left(\frac{2\mu}{3\mathfrak{K}T}\right)^{\frac{3}{2}} \tag{17}$$

which is essentially the equation given on replacing the bulk velocity in the usual accretion formula * by the mean thermal velocity of the atoms in the material.

Equation (17) shows that the condensation doubles its initial mass, $\mathscr{M}_0$ say, in a time

$$\frac{1}{8\pi\rho G^2 \mathscr{M}_0}\left(\frac{3\mathfrak{K}T}{2\mu}\right)^{\frac{3}{2}}$$

and increases in mass by a large factor in a time

$$\frac{1}{4\pi\rho G^2 \mathscr{M}_0}\left(\frac{3\mathfrak{K}T}{2\mu}\right)^{\frac{3}{2}}. \tag{18}$$

It may be noted that only a finite increase of mass can occur since ρ must decrease when the condensations have picked up an appreciable fraction of the diffuse material. A second effect is discussed below that also limits the increase of mass. The value of $\mathscr{M}_0$ is obtained by using the critical radius given by (16) and is

$$\mathscr{M}_0 = \frac{4\pi\sigma}{3}\left(\frac{9\mathfrak{K}T}{8\pi G\mu\sigma}\right)^{\frac{3}{2}}.$$

The time required for the condensation to attain large mass follows immediately by inserting this expression for $\mathscr{M}_0$ in (18). This gives

$$\left(\frac{\sigma}{6\pi\rho^2 G}\right)^{\frac{1}{2}} \tag{19}$$

which is of the same order as the time required for the direct condensation. Thus the time required for the whole process of planetary condensation is of order $(\sigma/\rho^2 G)^{\frac{1}{2}} \sim 10^8$ years.

There is an important question concerning the above result that now requires discussion. The gravitational acceleration due to the condensation acting on accreted material when near the capture radius given by (16) is

$$\frac{9k^2T^2}{4G\mathscr{M}\mu^2 m_H^2}. \tag{20}$$

For accretion to occur in the manner discussed above it is evidently necessary for this acceleration to exceed the differential acceleration across the capture radius (16) arising from the Sun's gravitational field. This requires

$$\frac{9k^2T^2}{4G\mathscr{M}\mu^2 m_H^2} > \frac{GM_\odot}{r^3} \cdot \frac{2G\mathscr{M}\mu m}{3kT}, \tag{21}$$

where r is the distance of the condensation from the Sun. This inequality gives an upper limit

$$M_\odot^{-\frac{1}{2}}\left(\frac{3kTr}{2G\mu m_H}\right)^{\frac{3}{2}} \tag{22}$$

for the mass to which the condensation can grow before modification in (17) becomes necessary. For $r = 7 \times 10^{13}$ cm., $\mu = 56$, $T = 273°$ K. the expression (22) is equal to $1{\cdot}1 \times 10^{28}$ gm. and the capture radius (16) corresponding to this mass is

* H. Bondi and F. Hoyle, *loc. cit.*

$1{\cdot}1 \times 10^{12}$ cm. This mass although smaller than the masses of the major planets is nevertheless very large compared with the mass of a condensation at the end of the direct condensation process. Thus for $T = 273°$ K., $\mu = 56$, $\sigma = 4$ gm. per cm.^{3} the expression for $\mathscr{M}_0$ gives $2{\cdot}1 \times 10^{23}$ gm.

The modification of (17) that becomes necessary when the mass exceeds $1{\cdot}1 \times 10^{28}$ gm. is easily discussed. By comparing the differential acceleration

$$\frac{2GM_\odot}{r^3} \cdot (\text{Accretion capture radius})$$

of the Sun's field with the gravitational acceleration

$$G\mathscr{M}/(\text{Accretion capture radius})^2$$

due to the condensation, it is seen that the accretion capture radius must be less than $r(\mathscr{M}/2M_\odot)^{\frac{1}{3}}$ in order for the disruptive effect of the Sun's field to be less than the attraction of the condensation. It appears therefore that

$$\frac{r}{2}\left(\frac{\mathscr{M}}{2M_\odot}\right)^{\frac{1}{3}} \tag{23}$$

is a reasonable value to take for the capture radius when $\mathscr{M} > 10^{28}$ gm. This value will be used in the following work. Equation (17) is modified to

$$\frac{d\mathscr{M}}{dt} \simeq \frac{\pi r^2 \rho}{4}\left(\frac{\mathscr{M}}{2M_\odot}\right)^{\frac{2}{3}}\left(\frac{3kT}{\mu m_H}\right)^{\frac{1}{2}} \tag{24}$$

which integrates immediately to give

$$\mathscr{M}^{\frac{1}{3}}(t_2) - \mathscr{M}^{\frac{1}{3}}(t_1) = \frac{\pi r^2 \rho (t_2 - t_1)}{12}\left(\frac{1}{2M_\odot}\right)^{\frac{2}{3}}\left(\frac{3kT}{\mu m_H}\right)^{\frac{1}{2}}, \tag{25}$$

where t_1, t_2 are two particular values of the time. Equation (25) gives the time required for accretion to increase the mass of a condensation from 10^{28} gm. at time t_1 up to a mass comparable with that of a major planet at time t_2. For example, we obtain

$$(t_2 - t_1) = 1{\cdot}4 \times 10^5 \text{ years}$$

on putting $\rho = 10^{-12}$ gm. per cm.^{3}, $r = 7 \times 10^{13}$ cm., $\mathscr{M}(t_1) = 10^{28}$ gm., and $\mathscr{M}(t_2) = 10^{30}$ gm. It is seen that the time required for the final stage in the growth of a planet is very short compared with the early stages of the condensation process. This conclusion is shown below to lead to a number of important consequences.

3.3. *The Effect of Collisons between the Condensations.**—The planetesimal hypothesis of Chamberlain and Moulton has been criticized by Jeffreys † who has shown that the planetesimals must volatilize through collisions instead of condensing into larger bodies. Although a similar evaporation process must occur in the present theory, there is the important difference that condensations are continually being re-formed by the process discussed in 3.1. Thus on the present theory a rough balance must be reached between evaporation and condensation. This balance will be maintained so long as all the condensations remain small, but the balance must evidently be disturbed if several condensations should attain masses large enough for rapid accretion to replace direct condensation.

* The writer is indebted to Professor W. H. McCrea for a valuable discussion concerning the work of the present subsection.

† H. Jeffreys, *M.N.*, **77**, 84, 1916.

The quantitative discussion of collisions is a problem of great complexity and will not be attempted in the present paper. There are, however, a number of qualitative features that are of importance, and these may be briefly mentioned.

(1) The approximate balance between evaporation and condensation is important because it prevents the whole of the gaseous material being condensed into small solid particles. This means that gaseous material must be available for the accretion process discussed in 3.2.

(2) The condensations discussed in 3.1 and 3.2 are seen to be highly exceptional since it has been assumed that these condensations do not suffer collisions. In particular it is assumed in 3.2 that the condensations move in a circular orbit about the Sun, and such an assumption could only occur in very special cases.

(3) In the early stages all condensations will be of approximately the same size, but as time proceeds there will be increasingly large fluctuations in the sizes. The largest condensations will be those that suffer least collisions. The approximate balance between evaporation and condensation will continue until a few condensations attain sufficiently large masses for direct condensation to be replaced by rapid accretion. The instability introduced by rapid accretion is very important, for it leads us to expect condensation into a few large planets rather than condensation into a swarm of small particles.

3.4. In concluding the present section it is interesting to note that equation (10), which can be written in the numerical form

$$\frac{ds}{dt} = 8{\cdot}66 \times 10^{-9} \qquad (10')$$

when $\rho = 10^{-12}$ gm. per cm.^{3} $\sigma = 4$ gm. per cm.^{3}, $T = 273°$ K., $f = 1$, $\mu = 56$, can be used to give a conclusive proof that the saturation vapour pressure of iron at 273° K. must be much less than the partial pressure existing in the diffuse planetary material. To see this, suppose for the moment that the situation were reversed, and the saturation vapour pressure exceeded the partial pressure. Then a solid body instead of increasing in size would evaporate. Moreover the rate of evaporation would exceed that given by reversing the sign of the right-hand side of (10′). Now this rate of evaporation is independent of the density of the gaseous material in which the condensation is placed, so that solid bodies in the solar system would be evaporating at the present time at a rate exceeding that given by reversing the sign of the right-hand side of (10′). If this were correct an iron meteorite in the solar system would evaporate completely in a period of a few years. For example, a spherical meteorite of radius 2 cm. maintained at a temperature of 273° K. would evaporate in less than eight years. This *reductio ad absurdum* supplies complete justification of the idea of direct condensation.

4. *The Rotations of the Planets.**—To obtain agreement with the observed motions of the planets it is necessary that the disk of diffuse material shall rotate in a direct sense about the Sun. It is by no means obvious that a condensation occurring in such a disk will have direct axial rotation, for although the angular momentum per unit mass of the diffuse material about the Sun increases outwards the orbital velocity on the other hand decreases outwards. The main object of the present section is to show that condensations attaining masses of the order of the

* The writer is indebted to Mr H. Bondi for valuable discussions on the subject-matter of this section, and also to Professor H. Jeffreys for suggesting an important improvement in the presentation.

major planets must have rapid axial rotation and that the rotation must be direct (the orbital motion about the Sun being direct).

The problem will be considered for the case of a condensation moving in a circular orbit about the Sun. This is appropriate for the condensations of largest mass, since condensations moving in circular orbits grow much more rapidly by accretion than condensations moving in orbits that are appreciably eccentric. The important feature of the circular orbit case is that the material undergoing accretion has an axis through the centre of the condensation and perpendicular to the plane of the disk as a principal axis of inertia. A further general simplifying feature is that gas pressure gradients in the diffuse material can be neglected in comparison with the gravitational field of the condensation. This follows because almost the whole of the angular momentum of the condensation is obtained during the final phase of the accretion process in which the mass is increased from 10^{28} gm. up to a value of order 10^{30} gm. The capture radius for accretion during this phase is of order

$$\frac{r}{2}\left(\frac{\mathscr{M}}{2M_\odot}\right)^{\frac{1}{3}},$$

so that the gravitational acceleration of the condensation acting on material near the capture radius is of order

$$2^{\frac{2}{3}}G\mathscr{M}^{\frac{1}{3}}M_\odot^{\frac{2}{3}}/r^2.$$

This expression may be compared with the acceleration arising from the gas pressure gradient, which must be of order

$$2^{\frac{1}{3}}\mathfrak{R}TM_\odot^{\frac{1}{3}}/r\mu\mathscr{M}^{\frac{1}{3}}.$$

The ratio of the acceleration due to the gas pressure gradient to the gravitational acceleration is seen to be of order

$$\mathfrak{R}Tr/2^{\frac{1}{3}}\mu GM_\odot^{\frac{1}{3}}\mathscr{M}^{\frac{1}{3}}, \tag{26}$$

which is small compared with unity during the final phase of accretion. For example, when $T=273°$ K., $r=7\times10^{13}$ cm., $\mu=56$, $\mathscr{M}=10^{30}$ gm. the ratio (26) is $1{\cdot}4\times10^{-2}$. It follows that the effect of gas pressure may be neglected during the final stages of the accretion process.

The following specific problem will now be considered. Let $V(r)$ be the orbital velocity of diffuse material at distance r from the Sun at time t', and let $V(r)$ satisfy the equation

$$V^2/r=M_\odot G/r^2. \tag{27}$$

That is, at time t' the material in the disk is everywhere moving in circular orbits and the centrifugal force balances solar gravity. Furthermore, let the centre of the condensation under consideration move in a circular orbit of constant radius r_0 (independent of time) with a constant angular velocity w given by

$$w=V(r_0)/r_0. \tag{28}$$

Our problem is to consider motion of material moving under the influence of the gravitational fields of the Sun and the condensation.

The above specification of the problem is slightly idealized in that the motion of the diffuse material at time t', as given by (27), could only occur if the gravitational field of the condensation were zero at times earlier than t'. This is not strictly correct, but the error introduced by this assumption is small since the main

contribution to the rotation of the condensation arises during the final stages of the accretion process. For example, if we regard t' as being the time when $\mathcal{M} = 10^{28}$ gm. the disturbance of the diffuse material produced by the condensation when $t < t'$ is small and can be neglected in comparison with the disturbances that arise during the time that the condensation grows from 10^{28} gm. to 10^{30} gm. It may be noted that taking the angular velocity of the centre of condensation about the Sun as constant in time means that second order terms are being neglected.

Take rectangular Cartesian axes with origin at the centre of the condensation. Let the x-axis be directed outwards along the line joining the Sun and the centre of the condensation, and let the z-axis be parallel to the axis of rotation of the disk. The y-axis is then chosen so as to complete a set of right-handed rectangular axes. The equations of motion of a particle moving under the influence of the Sun and the condensation are well known * and can be written as

$$\left.\begin{aligned} \ddot{x} - 2w\dot{y} - w^2 x &= \frac{\partial F}{\partial x}, \\ \ddot{y} + 2w\dot{x} - w^2 y &= \frac{\partial F}{\partial y}, \\ \ddot{z} &= \frac{\partial F}{\partial z}, \end{aligned}\right\} \tag{29}$$

where F can be written in second order as

$$\left.\begin{aligned} F &= \frac{G\mathcal{M}}{s} + \frac{M_\odot G}{r^3}\left(\frac{3}{2}x^2 - \frac{s^2}{2}\right), \\ s &= (x^2 + y^2 + z^2)^{\frac{1}{2}}. \end{aligned}\right\} \tag{30}$$

If (30) is used in (29) we obtain

$$\ddot{x} - 2w\dot{y} - w^2 x = -\frac{G\mathcal{M}x}{s^3} + \frac{2GM_\odot x}{r^3}, \tag{31}$$

$$\ddot{y} + 2w\dot{x} - w^2 y = -\frac{G\mathcal{M}y}{s^3} - \frac{GM_\odot y}{r^3}, \tag{32}$$

$$\ddot{z} = -\frac{G\mathcal{M}z}{s^3} - \frac{GM_\odot z}{r^3}. \tag{33}$$

It is important to notice that the present problem does not demand a complete solution of these equations, since we only require the angular momentum per unit mass $(x\dot{y} - y\dot{x})$ about the z-axis. The time dependence of this quantity can be obtained immediately from (31) and (32) which give

$$\frac{d}{dt}[(x\dot{y} - y\dot{x}) + w(x^2 + y^2)] = -\frac{3GM_\odot xy}{r^3}. \tag{34}$$

The term on the right-hand side of this equation disappears if we average with respect to all the accreted material, since the z-axis is a principle axis of inertia for the whole assembly of accreted particles. Thus on taking averages we can integrate immediately to obtain

$$\overline{(x\dot{y} - y\dot{x})}_t + w\overline{(x^2 + y^2)}_t = \overline{(xy - y\dot{x})}_{t'} + w\overline{(x^2 + y^2)}_{t'}.$$

* E. W. Brown, *Lunar Theory*, pp. 7, 19, Cambridge, 1896.

The second term on the left-hand side of this equation is evidently small compared with the other terms at the end of the accretion process. Thus if t is a sufficiently late stage in the accretion process we have

$$\overline{(xy-y\dot{x})}_t \simeq \overline{(x\dot{y}-y\dot{x})}_{t'} + w\overline{(x^2+y^2)}_{t'}. \tag{35}$$

Equation (35) gives the mean angular momentum per unit mass of material entering the condensation in terms of quantities that can be worked out from (27) taken together with the value of the accretion capture radius during the final stages of accretion.

The first step in working out the right-hand side of (35) is to obtain $x\dot{y}-y\dot{x}$ for a particle situated at time t' at the point (x,y,z). Consider an observer situated at the particular time t' at the centre of the condensation and let the observer remain in a fixed position relative to the Sun (since the observer possesses no orbital motion he can only be at the centre of the condensation at the particular times $(t'+2\pi m r_0/V(r_0))$, where $m=0, \pm 1, \pm 2, \ldots$). Let the observer use non-rotating rectangular axes that are coincident at time t' with the x,y,z axes. Then it is easy to show that the velocity of the particle situated at the point (x,y,z) at time t', as measured by the observer, is

$$\left[-wy, \quad V(r_0)+\left(\frac{\partial V}{\partial r}\right)_{r_0} x, \quad \dot{z}\right]$$

to the first order in small quantities. This may be compared with the velocity $(0, V(r_0), 0)$ of the centre of the condensation relative to the observer. Thus at time t' the particle appears to the observer to possess a velocity

$$\left[-wy, \quad \left(\frac{\partial V}{\partial r}\right)_{r_0} x, \quad \dot{z}\right] \tag{36}$$

relative to the centre of the condensation. The velocity of the particle in the (x,y,z) axes can be obtained from this result by noticing that the (x,y,z) axes are rotating about the z-axis with angular velocity w relative to the axes used by the observer. This rotation changes (36) to

$$\left[0, \quad \left\{\left(\frac{\partial V}{\partial r}\right)_{r_0} - w\right\} x, \quad \dot{z}\right] \tag{37}$$

and this is the velocity $(\dot{x},\dot{y},\dot{z})$ for the particle situated at the point (x,y,z) at time t'. Thus we have

$$(x\dot{y}-yx)_{t'} = x^2\left\{\left(\frac{\partial V}{\partial r}\right)_{r_0} - w\right\} \tag{38}$$

and (35) becomes *

$$\overline{(x\dot{y}-y\dot{x})}_t = \left(\frac{\partial V}{dr}\right)_{r_0} \overline{(x^2)}_{t'} + w\overline{(y^2)}_{t'}. \tag{39}$$

From (27) it follows that

$$\left(\frac{dV}{dr}\right)_{r_0} = -\frac{w}{2}$$

so that (39) can be written as

$$\overline{(x\dot{y}-y\dot{x})}_t = w\overline{(y^2-\tfrac{1}{2}x^2)}_{t'}. \tag{39'}$$

---

* A formula of this form is quoted in a footnote to a recent paper, *M.N.*, **105**, 312, 1945, for the case of a spherical stellar condensation occurring in a nebula.

The direction of axial rotation of the condensing material is in the same direction as the orbital motion if the left-hand side of (39′) is positive. Thus since it is to be expected that the distribution of accreted material will possess a greater extension in the tangential than in the radial direction, it follows that the axial rotation can be regarded as being direct when the orbital motion is direct.

If for simplicity the accretion process is taken to be spherically symmetric then it can be shown by averaging throughout a sphere of radius equal to the capture radius that

$$\overline{x^2}=\overline{y^2}=(\text{Accretion capture radius})^2/5 \simeq \frac{r_0^2}{20}\left(\frac{\mathcal{M}}{2M_\odot}\right)^{\frac{2}{3}}. \tag{40}$$

Thus (39′) gives

$$\overline{(x\dot{y}-y\dot{x})_i}=\frac{r_0V(r_0)}{40}\left(\frac{\mathcal{M}}{2M_\odot}\right)^{\frac{2}{3}} \tag{41}$$

which increases with $\mathcal{M}$. This variation can be allowed for by noticing that if $\mathcal{M}_f$ is the final mass of condensation then the final average angular momentum per unit mass of material in the condensation is given to sufficient accuracy by

$$\frac{r_0V(r_0)}{40\mathcal{M}_f}\frac{1}{(2M_\odot)^{\frac{2}{3}}}\int_0^{\mathcal{M}_f}\mathcal{M}^{\frac{2}{3}}d\mathcal{M}=\frac{3r_0V(r_0)}{200}\left(\frac{\mathcal{M}_f}{2M_\odot}\right)^{\frac{2}{3}}. \tag{42}$$

Now the average angular momentum per unit mass possessed by a spherical body of mass $\mathcal{M}_f$, density σ, and radius S, rotating uniformly in a period P is

$$\frac{4\pi S^2}{5P}=\frac{4\pi}{5P}\left(\frac{3\mathcal{M}_f}{4\pi\sigma}\right)^{\frac{2}{3}}. \tag{43}$$

By equating (42) and (43) we obtain the following expression for the period P of the condensation

$$P=\frac{160\pi}{3r_0V(r_0)}\left(\frac{3M_\odot}{2\pi\sigma}\right)^{\frac{2}{3}}=\frac{80(2\pi)^{\frac{1}{3}}M_\odot^{\frac{1}{3}}}{3^{\frac{1}{3}}\sigma^{\frac{1}{3}}G^{\frac{1}{2}}r_0^{\frac{1}{2}}}, \tag{44}$$

when $V(r_0)=(GM_\odot/r_0)^{\frac{1}{2}}$ is used. This equation gives the important result that the period of rotation is independent of the mass $\mathcal{M}_f$ of the planet. Moreover the period is rather insensitive to both σ and r_0.

An important question arises as to what value of σ we are to use in comparing (44) with observation. The major planets are so extensive that the mean density of (atmosphere + solid core) is only of order unity as opposed to the value of about 4 gm. per cm.^{3} that is likely to occur in the core. Thus if we wish to compare (44) with the observed periods of rotation it is necessary to use a value of order unity for σ.

A second point that may be noted is that in order to obtain a satisfactory theory of the origin of the minor planets and satellites * (see Section 5) it is necessary that the first planetary condensations shall be rotationally unstable. This means that the first condensations must have had appreciable shorter periods than the present periods of the major planets. Thus Lyttleton has estimated (using $\sigma=1$ gm. per cm.^{3}) that the rotation periods of the primitive planets must have been less than 7·2 hours, while periods as short as 4 hours have been discussed. It would appear therefore that a rotation period of about five hours for the first condensations satisfies the necessary requirements.

* R. A. Lyttleton, *M.N.*, **98**, 634, 1938.

When $\sigma = 1$ gm. per cm.^{3} and $r_0 = 7 \times 10^{13}$ cm. the period given by (44) is 4·67 hours which is in excellent agreement with the requirements discussed above. Moreover, since (42) is positive it follows that the axial rotation of condensation is in the same sense as its orbital motion about the Sun.

The present success of the theory should be regarded as an agreement between orders of magnitude rather than an agreement of exact values, since the calculated period depends on the value we have used for the accretion capture radius. This value for the capture radius is correct so far as an order of magnitude is concerned, but until an exact value is available the calculated period can evidently only be regarded as an order of magnitude estimate. Nevertheless the agreement obtained is both satisfactory and significant.

5. *The Major Planets, Terrestrial Planets and Satellites and the Asteroids.*— 5.1. It has been suggested by Lyttleton * that the terrestrial planets and satellites do not arise from separate condensation but from the rotational instability of primitive planets with masses comparable with those of the major planets. Lyttleton brings forward important evidence concerning both the satellite systems of the major planets and the differences of composition among the terrestrial planets in support of his theory. These arguments, although very significant, are circumstantial in character. That is, they concern the solar system in its present state and do not refer explicitly to the process of planetary condensation. Accordingly it is of interest to notice that the theory given in Sections 3 and 4 lends very powerful support to Lyttleton's theory. For in the first place it was shown in Section 3 that while a time of order 10^8 years is required for a condensation to attain a mass of order 10^{28} gm. the remaining phase of the condensation process in which mass rises from 10^{28} gm. to a value of order 10^{30} gm. takes only a time of order 10^5 years. Thus it is extremely improbable that a condensation could attain a mass of 10^{28} gm. without going on further to a mass of order 10^{30} gm. This means that a condensation will not in general stop acquiring new material when its mass becomes of the same order as the masses of the terrestrial planets but will quickly go on to attain a mass of the same order as the masses of the major planets. Indeed the only case in which a condensation would be likely to stop short at a mass of 10^{28} gm. would be if the diffuse gaseous material became exhausted when this mass was reached. This situation might account for a single instance of a body of mass 10^{28} gm. but could not be invoked to explain the large number of terrestrial planets and satellites that exist in the solar system. The second point in favour of Lyttleton's theory is that the rotation period of a condensation of mass 10^{30} gm. is probably less than the value required for rotational instability.

The above remarks show that a condensation attaining a mass of 10^{28} gm. will quickly go on to pick up a mass of the same order as the total mass of the diffuse material itself. The final stages of growth are so rapid (of order 10^5 years) that the terrestrial planets and satellites cannot be formed directly by this process. Moreover, not many condensations with masses of order 10^{30} gm. can occur. The reason for this is that the first permanent large condensations sweep up diffuse material at such a rapidly increasing rate that the supply of material is quickly exhausted. In this connection it is important to notice that accretion by large condensations has the effect of introducing hydrodynamic sinks into which material streams with a velocity comparable with the velocity of sound (of order

* R. A. Lyttleton, *loc. cit.*

$\frac{1}{2}$ km. per sec.). Indeed the only way for several large condensations to occur would be for their orbits to have widely separated radii. In this case the different condensations would tend to act as separate systems and would pick up material practically independently of each other. This could not occur for the case of two condensations in orbits of similar radii. These conclusions indicate that a few large planets will be formed with orbits of fairly widely separated radii, and that the type of rotational instability discussed by Lyttleton in his theory of the origin of the terrestrial planets and satellites probably occurs.

5.2. It was shown in Section 3 that

(i) The direct condensation process continues until a mass of about 10^{23} gm. is attained. The rate of direct condensation is approximately the same whether the orbit is circular or not.

(ii) For masses above 10^{23} gm. accretion of diffuse material becomes important when the orbit is approximately circular. A condensation in a circular orbit can attain a mass of order 10^{30} gm. before the mass of a condensation in an appreciably eccentric orbit rises much above 10^{23} gm.

The circular orbits referred to in (ii) must be maintained for a time of order 10^8 years (this being the time required for accretion to increase the mass from 10^{23} gm. to 10^{30} gm.). In view of the frequent collisions that occur between condensations (see 3.3) it is necessary to suppose that the chance of a condensation maintaining an approximately circular orbit for such a time is very small. This raises the important point that only a very small fraction of the bodies that attain a mass of order 10^{23} gm. by direct condensation can be regarded as growing by accretion to masses of planetary order. The situation is that a few exceptional condensations moving in practically circular orbits will increase more rapidly in mass and because of the rapidity of the final stages of accretion will exhaust the supply of diffuse material before the masses of the great majority of the condensations rise appreciably above 10^{23} gm.

The question naturally arises as to what has happened in the solar system to this large number of bodies of small mass. An indication of the nature of such a body is given by noting that a sphere of material of mean density 4 gm. per cm.^{3} and mass 10^{23} gm. has a radius of about 180 km. This value of the radius suggests that the large number of bodies of small mass are associated with the asteroids.

The number of asteroids known to occur in the solar system is large *, but the actual number is probably much greater still, since for observational reasons it becomes very difficult to detect such small bodies at distances of more than about 4 astronomical units. Thus very large numbers of undetected asteroids may well exist at distances from the Sun of the order of, or greater than, the radius of Jupiter's orbit.

A question arises concerning the rotations of the asteroids, for it can be shown that the process discussed in Section 4 cannot give rise to appreciable rotations for masses as small as 10^{23} gm. Appreciable rotations can arise, however, from the collisions discussed in 3.3. A sufficiently close encounter can endow a body with a rotational velocity of the order of, but less than, $(G\mathcal{M}/s)^{\frac{1}{2}}$, where $\mathcal{M}$, s are the mass and radius respectively of the body. If the rotational velocity were to exceed $(G\mathcal{M}/s)^{\frac{1}{2}}$ the body would evidently tend to break up. Such a break-up might occur

* Russell, Dugan and Stewart, *Astronomy*, I, p. 347, Boston, Mass., 1926.

through a direct collision. It follows that we may expect rotation periods of order

$$2\pi s^{\frac{3}{2}}/(G\mathscr{M})^{\frac{1}{2}}.$$

This expression is independent of s since $\mathscr{M} = 4\pi\sigma s^3/3$, which leads to a period of order

$$(3\pi/\sigma G)^{\frac{1}{2}}.$$

When $\sigma = 4$ gm. per cm.^{3} this gives a period of 1·65 hours.

Finally it is important to notice that there is no requirement in the present theory for all asteroids to have arisen in the manner described above. Thus it seems likely that bodies with masses of order 10^{23} gm. will also be formed in the following process:—

(i) By the same process of rotational instability that gives rise to the satellites and terrestrial planets.

(ii) In a head-on collision between two condensations with masses large compared with 10^{23} gm.

6. *The Physical State of the Planetary Condensations.*—Accreted material arrives at the surface of a condensation of radius s and mass $\mathscr{M}$ with kinetic energy of order $G\mathscr{M}/s$, and this kinetic energy is converted into heat when the material strikes the surface. For a condensation of mass 10^{28} gm. and density 4 gm. per cm.^{3} this energy is of order 10^{12} ergs. per gm., which is comparable with the energy required to vaporize one gram of iron. It is seen therefore that the heat generated by the infall of accreted material becomes large when $\mathscr{M}$ exceeds 10^{28} gm. This suggests that large planetary condensations will be heated until the solid phase is changed to a liquid or possibly gaseous phase. To decide this question it is necessary to compare the rate of supply of energy with the rate of loss of energy through radiation.

Consider a condensation that doubles its mass in a time t. Then the total rate of addition of mass is of order $\mathscr{M}/t$ per unit time, where $\mathscr{M}$ is the mass of the condensation at the time in question. Thus the rate at which accreted material strikes the surface of the condensation is $\mathscr{M}/4\pi s^2 t$ per unit surface area, and the material has energy of order $G\mathscr{M}/s$. It follows therefore that the rate at which kinetic energy is converted into heat at the surface of the condensation is of order

$$\left(\frac{\mathscr{M}}{4\pi s^2 t}\right)\cdot\left(\frac{G\mathscr{M}}{s}\right) = \frac{\mathscr{M}G\sigma}{3t} \text{ per unit surface area per unit time,}$$

since $4\pi s^3\sigma/3 = \mathscr{M}$. An estimate for the order of magnitude of this quantity can be obtained by using the value $t = 10^5$ years given in 3.2. Then for $\mathscr{M} = 10^{28}$ gm., $\sigma = 4$ gm. per cm.^{3} we have an energy addition of $3{\cdot}0 \times 10^8$ ergs per cm.^{3} per sec. to the surface of the condensation. Now it is to be expected that the temperature of the surface will be adjusted so that the rate of loss of energy by radiation is approximately equal to the rate of supply of energy due to the infall of accreted material. In the present example a surface temperature close to 2000 °K. is required for such a balance to be reached. This temperature is sufficiently high for masses of the order of the terrestrial planets to become liquid.

It will be noticed that the surface temperature is proportional to $\mathscr{M}^{\frac{1}{2}}$ and becomes of order 5000 °K. when $\mathscr{M} = 10^{30}$ gm. Thus masses of the order of the major planets probably become gaseous during the accretion process.

St. John's College,
Cambridge:
1946 *April* 12.

AN APPROXIMATE FORMULA FOR THE CONTINUOUS RADIATIVE ABSORPTION CROSS-SECTION OF THE LIGHTER NEUTRAL ATOMS AND POSITIVE AND NEGATIVE IONS

D. R. Bates

(Communicated by Professor H. S. W. Massey)
(Received 1946 June 25)

Summary

An approximate formula for the continuous radiative absorption cross-sections of the lighter neutral atoms and positive and negative ions is derived. It is used to discuss the general results to be anticipated and is applied specifically to *Be*, C^+, N^+, O^+, F^+, Ne^+ and Na^+.

---

1. *Introduction.*—The continuous radiative absorption cross-section can only be written in exact analytical form for hydrogen-like systems. Investigations by Kramers *, Sugiura †, Gaunt ‡, and Menzel and Pekeris §, have shown that for these it is simply

$$K_{\nu,(n\Sigma l)} = \frac{32\pi^2 e^6 R Z^4}{3\sqrt{3} c h^3 \nu^3 n^5} g, \qquad (1)$$

where ν is the frequency of the incident radiation (assumed of course to be above the threshold value), n and l are the principal and azimuthal quantum numbers of the initial state, Z is the nuclear charge (in atomic units), e, R, c and h have their standard significance, and g is a complicated factor, which however can be taken to be unity near the spectral head without much loss of accuracy. There has been a natural tendency to extend this result to other systems by merely replacing Z by some so-called effective nuclear charge, but it is now known that this procedure may lead to gross errors both in the magnitude and in the frequency variation, even in the case of a single electron outside a closed core, for which perhaps the approximation appeared most plausible (*cf.* Page ||, Bates ¶). It was pointed out by Bates ** that the failure arises essentially because it is necessary to use *different* values of Z to represent adequately the free and the bound wave functions of the active electron and the taking of any mean value in the determination of the transition integral is invalid as it cannot give the correct degree of overlap between the wave functions.

While, ideally at least, the cross-section can always be computed by employing purely numerical wave functions it is desirable to have available a convenient approximate formula of wider applicability than (1). Such a formula is derived in the present paper and by means of it *Be*, C^+, N^+, O^+, F^+, Ne^+ and Na^+ are treated.

* *Phil. Mag.*, **46**, 837, 1923.
† *J. Phys. Radium*, **8**, 113, 1927.
‡ *Phil. Trans.*, A, **229**, 163, 1929.
§ *M.N.*, **96**, 77, 1935.
|| *M.N.*, **99**, 385, 1939.
¶ *M.N.*, **106**, 432, 1946.
** *M.N.*, **100**, 25, 1939.

2. *Analysis.*—Regarding the transition as involving directly only a single electron described by suitably normalized initial and final wave function Ψ^i and Ψ^f the quantal expression for the cross-section is

$$K_\nu = \frac{32\pi^4 m^2 e^2}{3h^3 c} \cdot \frac{1}{\omega_i} \cdot \sum_i \sum_f v v \mathscr{C}_p \left| \int_\tau \Psi^* \, . \, \mathbf{r} \, . \, \Psi^f d\tau \right|^2, \tag{2}$$

where ω_i is the weight of the initial state, ν (as before) is the frequency of the incident radiation, v is the velocity of the ejected electron and $\mathscr{C}_p$ is a factor to allow for the distortion of the wave function of the passive electrons. Expressing Ψ^i and Ψ^f in the usual form using spherical coordinates r, θ, ϕ, and the customary notation,

$$\Psi^i = \left\{ \frac{2l+1}{4\pi} \cdot \frac{(l-m)!}{(l+m)!} \right\}^{\frac{1}{2}} R^i_{n,l}(r) \, P_l^m(\cos\theta) e^{im\phi}, \tag{3}$$

$$\Psi^f = \sum_{l=0}^{\infty} (2l+1) i^l e^{i\eta_{\epsilon,l}} R^f_{\epsilon,l}(r) \, P_l^0(\cos\theta), \quad (\epsilon = \tfrac{1}{2}mv^2) \tag{4}$$

and integrating over the angular coordinates formula (2) becomes

$$K_\nu = \sum_{\text{All states of ionization}} k_\nu, \tag{5}$$

$$k_\nu = \frac{32\pi^4 m^2 e^2}{3h^3 c} (v\nu) \left\{ 4\pi \mathscr{C}_p \left(C_{l-1} \left| \int_0^\infty R^i_{n,l} R^f_{\epsilon,l-1} r^3 dr \right|^2 + C_{l+1} \left| \int_0^\infty R^i_{n,l} R^f_{\epsilon,l+1} r^3 dr \right|^2 \right) \right\}, \tag{6}$$

C_{l-1} and C_{l+1} being constants tabulated by Bates.* Substituting numerically we have

$$k_\nu = 8{\cdot}5 \times 10^{-19} (I+\epsilon) \left\{ \mathscr{C}_p \epsilon^{\frac{1}{2}} \left(C_{l-1} \left| \int_0^\infty R^i_{n,l} R^f_{\epsilon,l-1} r^3 dr \right|^2 + C_{+1} \left| \int_0^\infty R^i_{n,l} R^f_{\epsilon,l+1} r^3 dr \right|^2 \right) \right\} \text{cm.}^2, \tag{7}$$

where now the ionization potential I, the ejection energy ϵ, and the nuclear distances r, are all expressed in Hartree units.†

As $\mathscr{C}_p$ is only slightly less than unity ‡ and is thus unimportant, it remains only to make suitable approximations for the radial wave functions and to carry out the integrations involved.

Following the procedure of Slater§ we may write,

$$R^i_{n,l} = \sum_t \sum_s a(s,t) r^t e^{-\alpha(s,t) r} \tag{8}$$

with

$t=0$ for $n=1$, $l=0$
$t=0,1$ $n=2$, $l=0$
$t=1$ $n=2$, $l=1$ etc.

In general a few values of s are sufficient. Great flexibility is obviously allowed. Any wave function in numerical form (obtained, say, by the self consistent field method) can be reproduced accurately by suitable choice of the various parameters $a(s,t)$ and $\alpha(s,t)$. Coulomb wave functions are included merely as a special case

* *M.N.*, **106**, 432, 1946.

† The Hartree units for energy and length are of course 13·53 e. volts and $5{\cdot}28_5 \times 10^{-9}$ cm. respectively.

‡ For example in the transition $O(^3P) \rightarrow O^+(^4S)$ D. R. Bates, R. A. Buckingham, H. S. W. Massey, and J. J. Unwin (*Proc. Roy. Soc.*, A, **170**, 322, 1939) found $\mathscr{C}_p$ to be 0·927.

§ *Phys. Rev.*, **42**, 32, 1939.

since for them

$$\left.\begin{aligned} R^i_{1,0} &= 2Z^{\frac{3}{2}}e^{-Zr}, \\ R^i_{2,0} &= \frac{1}{2\sqrt{2}}Z^{\frac{3}{2}}e^{-Zr/2}(Zr-2), \\ R^i_{2,1} &= \frac{1}{2\sqrt{6}}Z^{\frac{3}{2}}e^{-Zr/2}(Zr), \\ &\text{etc.} \end{aligned}\right\} \tag{9}$$

For the free wave functions of the lighter systems the best simple approximation is to neglect any deviation of the field of the residual core from its asymptotic form so that

$$R^f_{\epsilon,l} = e^{\pi z/2\epsilon^{\frac{1}{2}}} \cdot \frac{|\Gamma(l+1-iz\epsilon^{-\frac{1}{2}})|}{(2l+1)!}(2\epsilon^{\frac{1}{2}}r)^l e^{i\epsilon^{\frac{1}{2}}r} F(l+1+iz\epsilon^{-\frac{1}{2}}, 2l+2, -2i\epsilon^{\frac{1}{2}}r), \tag{10}$$ *

$F(a,b,cr)$ being the confluent hypergeometric series,

$$1 + \frac{a}{b\,.\,1}(cr) + \frac{a(a+1)}{b(b+1)\,.\,1.2}(cr)^2 + \ldots, \tag{11}$$

and z the excess of the nuclear over the electronic charge on the core (1, 2 and 0 for absorption by neutral atoms, positive ions and negative ions respectively). This procedure is satisfactory for orbitals with $l>0$ because their angular momentum limits the extent to which they penetrate the cloud of screening electrons around the nucleus. It has been shown to be justified for *He*, *O*, H^-, O^- and Na^- and there is little doubt that it is also justified for all other systems of neutral atoms, positive ions and negative ions up to at least the end of the second row of the Periodic Table. For orbitals with $l=0$ the approximation is certainly poor but this is not very serious as these are only involved in $p\rightarrow s$ transitions which are relatively slow compared with $p\rightarrow d$ transitions, at all frequencies for neutral atoms and positive ions, and except at low frequencies for negative ions.

Taking then the radial wave functions to be given by (8) and (10) it can be seen that the matrix elements depend essentially on integrals of the type

$$I = \int_0^\infty r^p e^{-qr} F(a,b,cr)dr. \tag{12}$$

These present no difficulty. Using the expansion (11) and integrating term by term † we have

$$I = \frac{p!}{q^{p+1}} F\left(a, p+1, b, \frac{c}{q}\right), \tag{13}$$

where $F\left(a, p+1, b, \frac{c}{q}\right)$ is the hypergeometric series

$$1 + \frac{(p+1)a}{b\,.\,1}\cdot\left(\frac{c}{q}\right) + \frac{(p+1)(p+2)a(a+1)}{b(b+1)\,.\,1.2}\left(\frac{c}{q}\right)^2 + \ldots. \tag{14}$$

Substituting in (7) at once yields the general formula for the absorption cross-section. This, however, besides being rather unwieldy is not of much interest. For the important cases, where the principal quantum number of the initial state is low, considerable simplification can be effected by using the recurrence

* W. Gordon, Z. *Physik*, **48**, 180, 1928.
† E. H. S. Burhop, *Proc. Roy. Soc.*, A, **148**, 272, 1935.

relation

$$p\left(1-\frac{c}{q}\right)F\left(a,p+1,b,\frac{c}{q}\right)=\left\{(p-a)\left(1-\frac{c}{q}\right)-(b-p-a)\right\}F\left(a,p,b,\frac{c}{q}\right)$$
$$-(p-b)F\left(a,p-1,b,\frac{c}{q}\right) \tag{15}$$

and observing that

$$F\left(a,b,b,\frac{c}{q}\right)=\left(1-\frac{c}{q}\right)^{-a}. \tag{16}$$

In particular with

$$n=1,\ l=0,\qquad k_\nu=8{\cdot}5\times10^{-19}(I+\epsilon)\,\mathscr{C}_p\{C_1|f^{(i)}(z,\epsilon)\sum_s a(s,0)g^{(i)}(z,\epsilon,\alpha(s,0))|^2\}\ \text{cm.}^2, \tag{17}$$

$$n=2,\ l=0,\qquad k_\nu=8{\cdot}5\times10^{-19}(I+\epsilon)\,\mathscr{C}_p\{C_1|f^{(i)}(z,\epsilon)\sum_s a(s,0)g^{(i)}(z,\epsilon,\alpha(s,0))$$
$$+f^{(ii)}(z,\epsilon)\sum_s a(s,1)g^{(ii)}(z,\epsilon,\alpha(s,1))|^2\}\text{cm.}^2, \tag{18}$$

$$n=2,\ l=1,\qquad k_\nu=8{\cdot}5\times10^{-19}(I+\epsilon)\,\mathscr{C}_p\{C_0|f^{(iii)}(z,\epsilon)\sum_s a(s,1)g^{(iii)}(z,\epsilon,\alpha(s,1))|^2$$
$$+C_2|f^{(iv)}(z,\epsilon)\sum_s a(s,1)g^{(iv)}(z,\epsilon,\alpha(s,1))|^2\}\ \text{cm.}^2, \tag{19}$$

where

$$2f^{(i)}(z,\epsilon)=f^{(ii)}(z,\epsilon)=8e^{\pi z/2\epsilon^{\frac{1}{2}}}|\,\Gamma(2+iz\epsilon^{-\frac{1}{2}})|\,\epsilon^{\frac{1}{4}}, \tag{20}$$

$$f^{(iii)}(z,\epsilon)=8e^{\pi z/2\epsilon^{\frac{1}{2}}}|\,\Gamma(1+iz\epsilon^{-\frac{1}{2}})|\,\epsilon^{\frac{1}{4}}, \tag{21}$$

$$f^{(iv)}(z,\epsilon)=8e^{\pi z/2\epsilon^{\frac{1}{2}}}|\,\Gamma(3+iz\epsilon^{-\frac{1}{2}})|\,\epsilon^{\frac{3}{4}}, \tag{22}$$

and

$$g^{(i)}(z,\epsilon,\alpha(s,0))=\frac{\{2\alpha(s,0)-z\}}{\{\alpha(s,0)^2+\epsilon\}^3}\,e^{-2\Theta(s,0)z\epsilon^{-\frac{1}{2}}}, \tag{23}$$

$$g^{(ii)}(z,\epsilon,\alpha(s,1))=\frac{\{(2\alpha(s,1)-z)(3\alpha(s,1)-z)-(\alpha(s,1)^2+\epsilon)\}}{\{\alpha(s,0)^2+\epsilon\}^2}\,e^{-2\Theta(s,1)z\epsilon^{-\frac{1}{2}}}, \tag{24}$$

$$g^{(iii)}(z,\epsilon,\alpha(s,1))=\frac{\{(\alpha(s,1)-z)(2\alpha(s,1)-z)(3\alpha(s,1)-z)-(3\alpha(s,1)-2z)(\alpha(s,1)^2+\epsilon)\}}{\{\alpha(s,1)^2+\epsilon\}^4}$$
$$e^{-2\Theta(s,1)z\epsilon^{-\frac{1}{2}}}, \tag{25}$$

$$g^{(iv)}(z,\epsilon,\alpha(s,1))=\frac{\{3\alpha(s,1)-z\}}{\{\alpha(s,1)^2+\epsilon\}^4}\,e^{-2\Theta(s,1)z\epsilon^{-\frac{1}{2}}}, \tag{26}$$

$$\Theta(s,t)\text{ being arc tan }(\epsilon^{\frac{1}{2}}/\alpha(s,t)). \tag{27}$$

The f-functions which of course do not depend on the parameters $a(s,t)$, $\alpha(s,t)$, are given numerically in Table I for $z=1$ (neutral atoms), $z=2$ (positive ions) and $z=0$ (negative ions). Computation of the g-functions in any individual case is rapid as, apart from the simple exponential factor *, they are all of the form

$$\frac{(A+B\epsilon)}{(C+\epsilon)^4}, \tag{28}$$

where A, B and C are constants.

* It will be noted that for $z=0$ (negative ions) this disappears.

TABLE I

ϵ (units 13·53 e. volts)	0·0	0·2	0·4	0·6	0·8	1·0
$2f^{(i)}(1,\epsilon)=f^{(ii)}(1,\epsilon)$	20·1	22·0	23·7	25·4	26·9	28·4
$2f^{(i)}(2,\epsilon)=f^{(ii)}(2,\epsilon)$	56·7	58·1	59·5	60·8	62·1	63·4
$2f^{(i)}(0,\epsilon)=f^{(ii)}(0,\epsilon)=8\epsilon^{\frac{3}{2}}$	0	2·39	4·02	5·45	6·77	8·00
$f^{(iii)}(1,\epsilon)$	20·1	20·1	20·1	20·1	20·1	20·1
$f^{(iii)}(2,\epsilon)$	28·4	28·4	28·4	28·4	28·4	28·4
$f^{(iii)}(0,\epsilon)=8\epsilon^{\frac{3}{2}}$	0	5·35	6·36	7·04	7·57	8·00
$f^{(iv)}(1,\epsilon)$	20·1	29·5	38·3	46·8	55·1	63·4
$f^{(iv)}(2,\epsilon)$	113	127	141	154	167	179
$f^{(iv)}(0,\epsilon)=16\epsilon^{\frac{5}{2}}$	0	2·14	5·09	8·45	12·1	16·0

3. *Discussion.*—Inspection of the absorption cross-section formula reveals some general results of interest. The main features are discussed most readily if only a single term $\alpha(s,t)r^t e^{-\alpha(s,t)r}$ in expression (8) is taken.

(*a*) *Magnitude.*—Bearing in mind that owing to the normalization * the values of $a(s,t)$ are not entirely independent of those of $\alpha(s,t)$, it can be seen that the cross-section of positive ions tends to be rather smaller, while the cross sectionof negative ions tends to be larger than that of the corresponding neutral atoms—the change in the g-function more than compensating for the change in the f-function. Similarly the cross-section for systems with incomplete shells tends to be greater than that for systems with filled shells (apart of course from the weighting factors). Exceptions to these rules can however occur due to the cancellation phenomena (*v. infra*).

(*b*) *Frequency Variations.*—At high frequencies the cross-section must always approach zero. If A and B [in (28)] are of the same sign the rate at which the fall-off occurs depends mainly on the magnitude of C—that is of $\alpha(s,t)$. For diffuse systems $\alpha(s,t)$ is small and consequently the rate of fall-off is fast while for compact systems $\alpha(s,t)$ is large and consequently the rate of fall-off is slow—indeed the decrement of the g-function may not become dominant until high frequencies are reached and near the spectral head the cross-section may actually rise with frequency due to the growth of the f-function and of the external factor $(I+\epsilon)$. In special circumstances (the conditions for which are similar to those described below in connection with the cancellation phemonena) A may be abnormally small and this also may lead to a cross-section initially increasing with frequency.

A more complex situation arises if A and B are not of the same sign. The contribution to the cross-section from a matrix element having this property may decrease more rapidly than usual and passes through a zero minimum † at

$$\epsilon_c = -A/B. \tag{29}$$

Such cancellation is of great importance in detailed computations since in the region of the minimum the cross-section is extremely low and is so sensitive to the

* If the wave function is normalized to unity then $a(s,t)=(2\alpha(s,t))^{t+\frac{3}{2}}\,((2t+2)!)^{-\frac{1}{2}}$. (But see Section 4.)

† It may be pointed out however that a full investigation including allowance for the non-separability of the wave functions shows the minimum to be zero only to the first approximation. In addition, of course, transitions from states other than with $l=0$ involves two matrix elements. These are unlikely to cancel at exactly the same frequency.

exact form of the wave functions employed that accurate evaluation is very difficult. It is of interest to note that if the hydrionic wave functions (9) are used with $Z=z$ (the early approximation mentioned in Section 1) cancellation is *never* possible, as can be verified in a few special cases by substituting

$$\left.\begin{aligned}&\alpha(s,0)=Z=z \qquad (\text{when } n=1,\ l=0),\\ &\alpha(s,0)=\alpha(s,1)=\tfrac{1}{2}Z=\tfrac{1}{2}z \quad (\text{when } n=2,\ l=0 \text{ or } 1)\end{aligned}\right\} \tag{30}$$

in (23), (24), (25) and (26). This erroneous result is illustrative of the defects already emphasized. Taking now $Z\neq z$ and again referring to (23), (24), (25) and (26) it is evident that for $l=0\rightarrow l=1$ transitions cancellation does not occur if in the initial state $n=1$, but that it may occur in $n=2$. Neglecting the small effect of the extreme inner loop of the bound wave function the condition is simply

$$\epsilon_c=5\alpha(s,1)^2-5\alpha(s,1)z+z^2\geqslant 0. \tag{31}$$

This requires

$$\alpha(s,1)\leqslant 0{\cdot}28z, \quad \text{i.e. } Z\leqslant 0{\cdot}56z, \tag{32}$$

or

$$\alpha(s,1)\geqslant 0{\cdot}72z, \quad \text{i.e. } Z\geqslant 1{\cdot}44z. \tag{33}$$

Clearly (32) is not satisfied by any system and can be ignored. (33) however, is of great significance. It is obviously satisfied for negative ions since for these $z=0$. Cancellation therefore is to be anticipated—though in most instances the value of ϵ_c is probably such that it occurs beyond the range normally of importance. Examination of the numerically determined wave functions of the neutral atoms shows that for lithium (33) is not satisfied but that for the more compact beryllium it is satisfied. The cross-section for the latter should therefore fall off more rapidly than that for the former and should pass through a zero minimum. While no calculations have been done on the relevant wave functions of the positive ions it is evident that the critical ratio Z/z is less than for the corresponding neutral atoms. Hence Be^+ should not comply with (33). Though it is not possible to make a definite prediction regarding B^+, ϵ_c, whether positive or negative, is likely to be small and a high degree of cancellation may occur in the interesting region near the spectral head. In highly ionized systems (33) is not satisfied. Similar predictions to the above can be arrived at for $n>2$. They are in good agreement with such individual calculations as have been done—thus, for example, in both K (Bates *) and Ca^+ (Bates and Massey †) there is considerable cancellation at low frequencies, ϵ_c being small and positive for the former and small and negative for the latter. Little can be said about the relatively unimportant $l=1\rightarrow l=0$ transitions as for these the approximation employed is not sufficiently good. Considering finally the $l=1\rightarrow l=2$ transitions it can be seen that if $n=2$ cancellation cannot occur.‡ It does, however, become possible for higher values of n. The condition for $n=3$ can be shown to be

$$\epsilon_c=\tfrac{1}{3}\{21\alpha(s,2)^2-14\alpha(s,2)+2z^2)\}\geqslant 0. \tag{34}$$

Thus

$$\alpha(s,2)\leqslant 0{\cdot}21z, \quad \text{i.e. } Z\leqslant 0{\cdot}63z \tag{35}$$

---

* *Proc. Roy. Soc.*, A, **188**, 350, 1947.
† *Proc. Roy. Soc.*, A, **177**, 281, 1941.
‡ An analogous case to $l=0\rightarrow l=1$ transitions with $n=1$.

or

$$\alpha(s,2) \geqslant 0{\cdot}46z, \quad \text{i.e. } Z \geqslant 1{\cdot}38z. \tag{36}$$

For negative ions and neutral atoms ϵ_c is certainly positive but is probably so large that serious cancellations will only occur at very high frequencies. The decrease of Z/z in going to positive ions may however make the effect important for them in at least some instances.

4. *Application.*

4.1. *Neutral Atoms.*—Calculations of the absorption cross-section of all the neutral atoms from *H* to *Na* save *Be* have already been carried out (*cf.* review by Bates *). If only for the sake of completeness it was thought worthwhile to apply the formula developed in the present paper to this exception. The various parameters involved in (8) were choosen so as to yield a satisfactory representation of the bound wave function as determined by Hartree and Hartree.† It was found that

$$\left.\begin{aligned} a(0,0) &= 1{\cdot}58, & \alpha(0,0) &= 2{\cdot}14, \\ a(1,0) &= 1{\cdot}72, & \alpha(1,0) &= 6{\cdot}50, \\ a(0,1) &= -0{\cdot}62, & \alpha(0,1) &= 0{\cdot}87, \\ a(1,1) &= -1{\cdot}10, & \alpha(1,1) &= 1{\cdot}46. \end{aligned}\right\} \tag{37}$$

The *g*-functions can be computed by substituting (37) in (23) and (24) (with $z=1$). Table I gives the *f*-functions. For *Be* the constant $C_1=2$ and the factor $\mathscr{C}_p \simeq 1$. The cross-section can thus be obtained at once from (8). It is tabulated below. There is a zero minimum due to cancellation at a value of ϵ just below 1·0. The results are typical of such a system (Section 3).

TABLE II

Absorption cross-section for Be

Energy of ejected electron (ϵ) (units 13·53 e. volts).	0	0·1	0·2	0·3	0·4	0·5
Cross-section *Be* (k_ν) (Ionization potential 9·28 e. volts). (units 10^{-18} cm.^{2})	8·2	4·0	2·0	1·0	$0{\cdot}5_0$	$0{\cdot}2_4$

4.2. *Positive Ions.*—As little attention has hitherto been paid to absorption by positive ions, a number of cases are treated here. Bound wave functions are available for O^+ (Hartree, Hartree and Swirles ‡) and Na^+ (Hartree and Hartree §). Using them as a basis we have in the case of O^+

$$\left.\begin{aligned} a(0,1) &= 3{\cdot}00, & \alpha(0,1) &= 1{\cdot}82, \\ a(1,1) &= 10{\cdot}70, & \alpha(1,1) &= 3{\cdot}28, \end{aligned}\right\} \tag{38}$$

and in the case of Na^+

$$\left.\begin{aligned} a(0,1) &= 4{\cdot}32, & \alpha(0,1) &= 2{\cdot}23, \\ a(1,1) &= 24{\cdot}80, & \alpha(1,1) &= 4{\cdot}30. \end{aligned}\right\} \tag{39}$$

---

* *M.N.*, **106**, 432, 1946.
† *Proc. Roy. Soc.*, A, **150**, 9, 1935.
‡ *Phil. Trans.*, A, **238**, 229, 1939.
§ Kindly supplied prior to publication by Professor D. R. Hartree.

Further, $C_0=1$, $C_2=2$ (O^+) and $C_0=2$, $C_2=4$ (Na^+). Taking $\mathscr{C}_p \simeq 1$ the cross-sections can be evaluated from (25) and (26) (with $z=2$), Table I and (19). The results are given in Table III.

TABLE III

Absorption cross-section of O^+ and Na^+

Energy of ejected electron (ϵ) (units 13·53 e. volts).	0	0·2	0·4	0·6	0·8	1·0
Cross-section O^+ (k_ν) (Ionization potential 34·94 e volts).	8·1	7·8	7·3	6·8	6·4	6·1
Cross-section Na^+ (k_ν) (Ionization potential 47·14 e. volts).	7·1	7·3	7·4	7·4	7·4	7·3

(units 10^{-18} cm.^{2})

An important feature is the slowness of the fall-off of the cross-section with frequency—the contrast with the behaviour for *Be* is very striking.

Bound wave functions for none of the other positive ions with outer $2p$ shells appear to have been computed. Nevertheless it is possible to make a fair preliminary estimation of the cross-sections. The O^+ and Na^+ investigations showed that the main contribution to the transition integral comes from comparatively large radial distances. This enables one of the two sets of parameters normally employed in describing the bound wave function to be omitted so that we have simply

$$R^i_{2,1} = are^{-\alpha r} = a_N(2\alpha^{\frac{5}{2}}/\sqrt{3})re^{-\alpha r}. \tag{40}$$

It was found that a_N and α could be chosen * so that both the magnitude and the frequency variation of the cross-section of O^+ and Na^+ as given by the more precise method could be accurately reproduced. The actual values are as follows:

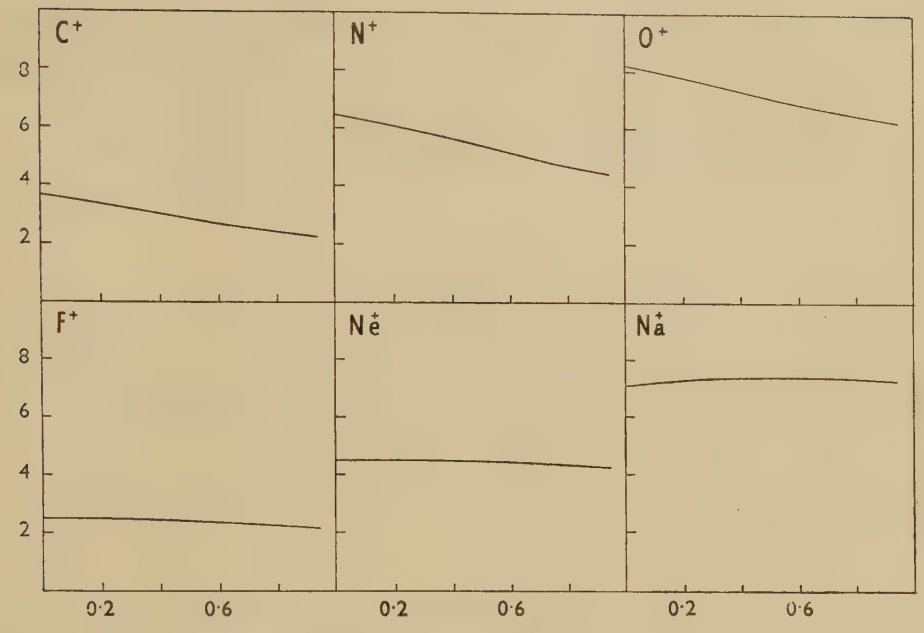
$$O^+ \quad a_N=0{\cdot}82 \quad \alpha=2{\cdot}00,$$
$$Na^+ \quad a_N=0{\cdot}69 \quad \alpha=2{\cdot}42.$$

It seems reasonable to obtain the parameters for the other positive ions of the same type (C^+, N^+, F^+, Ne^+) by interpolation or extrapolation. This procedure is unlikely to lead to undue error as the total range involved is not great. The diagram depicts the cross-sections finally derived. The irregularity to be observed in passing along the group is due principally to the constants C_0 and C_2 which have the values given below.

	C_0	C_2
C^+	1/3	2/3
N^+	2/3	4/3
O^+	1	2
F^+	4/9	8/9
Ne^+	1	2
Na^+	2	4

In all cases it is assumed that the resulting doubly charged ion is in the ground state. For F^+ and Ne^+ transitions to the excited states with the same configuration

* The wave functions obtained are not of course normalized to unity (which requires $a_N=1$) since they are inaccurate in (for the present purpose) unimportant region of high electron density.



Absorption cross-section of the positive ions C^+, N^+, O^+, F^+, Ne^+, Na^+.
Ordinates: absorption cross-section in units 10^{-18} cm.^{2}.
Abscissae: energy of ejected electron in units I_H (13·53 e.v.).

		e.v.		e.v.
Ionization potential:	C^+	24·26	F^+	34·91
	N^+	29·44	Ne^+	40·86
	O^+	34·94	Na^+	47·14

can also occur. The associated cross-sections can be obtained very simply from the data given, since the transition integrals (for a particular ϵ) are unaltered. Different values of C_0 and C_2 have of course to be used (the ratio to those for the ground state being 5/4 for $F^{++}(^2D)$, 3/4 for $F^{++}(^2P)$, 5/9 for $Ne^{++}(^1D)$ and 1/9 for $Ne^{++}(^1S)$, and allowance has to be made for the change in I (the ionization potential).

4.3. *Negative Ions.*—Semi-analytical methods have already been applied to H^- * and to Na^-.† The technique is of limited use since, as mentioned in Section 2, in the important group of negative ions where $l=1$ in the bound state, transitions to the free orbitals with $l=0$ predominate at low energies and for these the simple approximation is very poor.

As a fair number of representative negative ions have been investigated and as, owing to the poorly defined bound state, the accuracy obtainable at present is low, further detailed calculations were not thought to be justified.

Department of Mathematics, University College,
London, W.C.1:
1946 *June* 23.

* H. S. W. Massey and D. R. Bates, *Ap. J.*, **91**, 202, 1940.
† K. Huang, *Ap. J.*, **101**, 196, 1945.

CALCULATION OF THE CROSS-SECTION OF NEUTRAL ATOMS AND POSITIVE AND NEGATIVE IONS TOWARDS THE ABSORPTION OF RADIATION IN THE CONTINUUM

D. R. Bates

(Communicated by Professor H. S. W. Massey)

(Received 1946 May 25)

Summary

The use of quantum mechanics in the calculation of the cross-sections of neutral atoms and of positive and negative ions towards the absorption radiation in the continuum is reviewed with particular reference to the practical difficulties and limitations on the accuracy obtainable. The general types of results to be expected are discussed. The detailed calculations that have already been carried out are summarized and the possibilities of future work are indicated.

1. *Introduction.*—The experimental determination of the continuous radiative absorption cross-section of neutral atoms and positive and negative ions presents great technical difficulties. Though these are gradually being overcome by such work as that of Ditchburn, there appears little hope of a comprehensive series of measurements in the near future. In view of the important astrophysical applications it is therefore desirable to consider how far our knowledge is likely to be increased by theory. A general survey is rendered possible by the insight gained from the detailed calculations which have been carried out in recent years.

2. *The Nature of the Problem.*—If an $\mathcal{N}$-electron system is described in its initial state (weight ω_i) by a wave function $\Psi^i(\boldsymbol{r}_1 - \boldsymbol{r}_N)$ normalized to represent a unit cloud and in its final state (weight not explicitly involved) by a wave function $\Psi^f(\boldsymbol{r}_1 - \boldsymbol{r}_N)$, normalized to represent a unit core together with an outgoing wave of unit amplitude, then the absorption cross-section* K_ν is given by the usual quantal formula †

$$K_\nu = \frac{32\pi^4 m^2 e^2}{3h^3 c} \cdot \frac{1}{\omega_i} \sum_{\substack{\text{Initial wave} \\ \text{functions}}} \sum_{\substack{\text{Final wave} \\ \text{functions}}} (\nu v) \left| \int_\tau \Psi^{*i} (\Sigma \boldsymbol{r}) \Psi^f d\tau \right|^2, \qquad (1)$$

where ν is the frequency of the incident radiation and v is the velocity of the ejected electron (which are of course connected by the relation

$$h\nu = E_0 + \tfrac{1}{2} m v^2, \qquad (2)$$

E_0 being the limiting energy). Standard notation is employed for the electronic mass, charge etc. The exact application of this would certainly yield accurate results. Unfortunately such a procedure is usually impracticable and resort has to be made to approximations.

* The cross-section K_ν is such that the intensity of a beam of light of frequency ν is reduced by a factor e^{NK_ν} in passing through 1 cm. length of a gas containing N atoms (or positive or negative ions) per cm.^{3}.

† In addition to being expressed in terms of dipole length (as here) the cross-section can also be expressed in terms of dipole velocity or acceleration. These alternatives have not been much used (but see Section 4.3.1).

For many electron systems the wave functions of the individual orbitals can satisfactorily be separated and (using spherical coordinates r, θ, ϕ) written in the form

$$\left(\frac{(2l+1)}{4\pi}\cdot\frac{(l-m)!}{(l+m)!}\right)^{\frac{1}{2}} R_{n,l}(r)\,P_l^m(\cos\theta)e^{im\phi} \tag{3}$$

if bound, and in the form

$$\sum_{l=0}^{\infty}(2l+1)i^l e^{i\eta_{\epsilon,l}} R_{\epsilon,l}(r)P_l^0(\cos\theta) \tag{4}$$

if free. Here as usual n is the principal quantum number, l is the azimuthal quantum number, m is the magnetic quantum number, and ϵ is the energy of the ejected electron. The normalization of the radial wave functions for bound orbitals is such that

$$\int_0^\infty R^2_{n,l}r^2dr = 1 \tag{5}$$

and for free orbitals it is such that

$$R_{\epsilon,l}\sim\frac{1}{kr}\sin(kr-\tfrac{1}{2}l\pi+\eta_{\epsilon,l}+\alpha\log 2kr), \tag{6}$$

where

$$k=\frac{2\pi mv}{h},$$

$$\alpha=\frac{2\pi e}{hv}(ze)$$

and ze is the excess of the positive over the negative charge on the residual ion or atom.* After integration over the angular coordinates the original formula (1) for the absorption cross-section reduces to

$$K_\nu=\sum_{\substack{\text{All states of}\\ \text{ionization}}} k_\nu, \tag{7}$$

$$k_\nu=\frac{32\pi^4m^2e^2}{3h^3c}(\nu v)\prod_{\substack{\text{Passive}\\ \text{electrons}}}\left|\int_0^\infty R^i_{n,l}R^f_{n,l}r^2dr\right|^2\cdot$$

$$\underset{\substack{\text{Active}\\ \text{electron}}}{4\pi}\left\{C_{l-1}\left|\int_0^\infty R^i_{n,l}R^f_{\epsilon,l-1}r^3dr\right|^2+C_{l+1}\left|\int_0^\infty R^i_{n,l}R^f_{\epsilon,l+1}r^3dr\right|^2\right\}, \tag{8}$$

In general, of the partial cross-sections involved in the expansion (7), the only ones of interest are those relating to the lower states of ionization. The analytical expression for the constants C_{l-1} and C_{l+1} is cumbersome and their actual values are tabulated below in the more important cases. (Table I.) It is apparent that the selection rules for ionization are simply:

(i) the multiplicity can only change by ± 1;

(ii) the total angular momentum L can only change by $0, \pm 1, \ldots \pm 1$ (the initial azimuthal quantum number of the ejected electron).

* The method of carrying out this normalization for non-zero ze is discussed by J. Hargreaves (*Proc. Camb. phil. Soc.*, **25**, 75, 1928), that for zero ze is trivial.

It can be seen, therefore, that the calculation of the cross-section depends essentially on computing accurately integrals of the type

$$\int_0^\infty R^i_{n,\,l} R^f_{n,\,l} r^2 dr \quad \text{(passive electrons)}, \tag{9}$$

$$\int_0^\infty R^i_{n,\,l} R^f_{\varepsilon,\,\begin{cases} l-1 \\ l+1 \end{cases}} \text{or } r^3 dr \quad \text{(active electron)}. \tag{10}$$

The first give the correction to the cross-section necessary to allow for the distortion of the wave functions of the passive electrons due to the transition of the active electron. As this distortion is usually only slight the values of the integrals are very little less than unity and can be determined with adequate precision even using a comparatively crude model. The second, however, form the fundamental part of the matrix element and have to be treated with the greatest care. Various approximations have been used for the relevant bound and free wave functions; the earlier ones were generally based on simple self consistent fields*, but recently consideration has been given to such effects as exchange, polarization and non-separability.† Sufficient work has not yet been done to indicate the relative influence of these refinements on the wave functions of the different systems of the Periodic Table.

The situation is complicated by the fact that under certain circumstances the integral itself becomes extremely sensitive to the exact form of the wave functions owing to the contributions from the positive and negative portions of the integrand almost completely cancelling (*cf.* Bates ‡). When there is a possibility of this, a reliable calculation of the cross-section may, indeed, not be practicable.

The incidence of such cancellation will be considered later when discussing the various individual systems. It is sufficient to mention here that the phenomenon is probably quite common for atoms and positive ions, but does not occur for negative ions (at least near the limiting energy), since for these, owing to the rapid decrease of the attractive field, the vital node between the first and second regions of large density of the free wave function is so far from the nucleus that the density of the bound wave function is there very small. As a consequence the contributions to the transition integral from the two regions are not of comparable magnitude.

Calculations on the cross-section of negative ions may encounter, however, a special difficulty of an entirely different type, as was pointed out by Bates and Massey.§ The undistorted but polarized field of the core may be close to that required to make the free wave function of zero energy equivalent to a bound wave function. In such a case the cross-section is so sensitive to small changes in the field that for low energies the calculated value may be considerably in error. This sensitivity (which may loosely be regarded as a resonance effect) arises physically because, at exact resonance, the free wave function gives an abnormally great density in the region of the nucleus—for a given solution at the origin the

* J. Hargreaves, *Proc. Camb. phil. Soc.*, **25**, 75, 1928 and B. Trumpy, *Z. Phys.*, **71**, 720, 1931.

† H. S. W. Massey and D. R. Bates, *Ap. J.*, **91**, 202, 1940 ; D. R. Bates and H. S. W. Massey, *Proc. Roy. Soc.*, A, **177**, 281, 1941 ; D. R. Bates and H. S. W. Massey, *Phil. Trans.*, A, **239**, 269, 1943 ; K. Huang, *Ap. J.*, **101**, 196, 1945 ; D. R. Bates, *Proc. Roy. Soc.*, A, **188**, 350, 1947.

‡ *Proc. Roy. Soc.*, **188**, 350, 1947.

§ *Phil. Trans.*, A, **239**, 269, 1943.

TABLE I

Initial state	Final state	Value of C_{l-1}	Value of C_{l+}
$s\ {}^2S$	$1S$	...	1
$s^2\ {}^1S$	$s\ {}^2S$	...	2
$p\ {}^2P$	$1S$	1/3	2/3
$p^2\ {}^3P$	$p\ {}^2P$	2/3	4/3
$p^2\ {}^1D$	$p\ {}^2P$	2/3	4/3
$p^2\ {}^1S$	$p\ {}^2P$	2/3	4/3
$p^3\ {}^4S$	$p^2\ {}^3P$	1	2
	$p^2\ {}^1D$	0	0
	$p^2\ {}^1S$	0	0
$p^3\ {}^2D$	$p^2\ {}^3P$	1/2	1
	$p^2\ {}^1D$	1/2	1
	$p^2\ {}^1S$	0	0
$p^3\ {}^2P$	$p^2\ {}^3P$	1/2	1
	$p^2\ {}^1D$	5/18	5/9
	$p^2\ {}^1S$	2/9	4/9
$p^4\ {}^3P$	$p^3\ {}^4S$	4/9	8/9
	$p^3\ {}^2D$	5/9	10/9
	$p^3\ {}^2P$	1/3	2/3
$p^4\ {}^1D$	$p^3\ {}^4S$	0	0
	$p^3\ {}^2D$	1	2
	$p^3\ {}^2P$	1/3	2/3
$p^4\ {}^1S$	$p^3\ {}^4S$	0	0
	$p^3\ {}^2D$	0	0
	$p^3\ {}^2P$	4/3	8/3
$p^5\ {}^2P$	$p^4\ {}^3P$	1	2
	$p^4\ {}^1D$	5/9	10/9
	$p^4\ {}^1S$	1/9	2/9
$p^6\ {}^1S$	$p^5\ {}^2P$	2	4

asymptotic amplitude of the function $\{(kr)R^f_{\epsilon,\ l-1 \text{ or } l+1}\}$ goes through zero if the field is varied through its critical value.

Before leaving the general subject of negative ion cross-section calculations, it may be emphasized that much basic experimental and theoretical work has to be done before proper progress can be made. Thus there is still doubt as to the stability of some negative ions and for only a few is the limiting (that is detachment) energy known reliably. A suitable technique has not yet been developed for obtaining accurate bound wave functions—particularly at large radial distances.* In certain instances, even the configuration of the ground state is undecided. The information concerning excited states is meagre.

* Though this problem has not been entirely solved for neutral atoms and positive ions the difficulties are less acute.

3. *The Form and Magnitude of the Cross-section Frequency Curves.*—For neutral atoms and positive ions the cross-section has a finite and (ignoring exact cancellation) non-zero value at the spectral head. In the hydrogen-like case it varies approximately as the inverse cube of the frequency of the incident radiation (Section 4.1.1). This result, however, is certainly not general; the assumption that it is leads to gross errors.

The different possibilities can readily be understood by considering a simplified model valid, at least for the lighter systems.* Neglecting the unimportant nodes close to the nucleus the bound wave function can be treated as giving a single major region of high density and the significant part of the free wave function as giving two major regions of high density on either side of a node X. The transition matrix element may then be written in the form $(A-B)$, A being the contribution from within X and B being the contribution from without X. It is convenient also to adjust the normalization to eliminate the multiplying factor (νv) from the expression for the cross-section. Now, as the frequency (and hence the energy of the ejected electron) is increased, X moves towards the nucleus. If the bound wave function is relatively compact, both A and B tend to become larger, but the change in A is the more important (B remaining in fact very small). In consequence the cross-section rises with frequency (example: neon†). For relatively diffuse bound wave functions, however, A may decrease instead of increasing and B may become appreciable so that the cross-section falls with frequency (example: boron†). Should A be only slightly greater than B, this fall is very rapid (example: sodium‡) and indeed in some cases the cross-section actually reaches a zero minimum due to $(A-B)$ changing sign§ (example: potassium‖). Finally, if A is initially slightly less than B, there may once again be a rise with frequency (example: singly ionized calcium¶). Though this survey is not exhaustive, it shows the main trends. The introduction of more detailed wave functions and the combination of the $l\rightarrow l+1$ and $l\rightarrow l-1$ transitions only leads to minor variants.

In marked contrast to that of neutral atoms and positive ions, the cross-section of negative ions normally tends to zero at the spectral head. This arises because the field does not build up the density near the nucleus of the free orbitals of low energy. The matter has been discussed by Massey**. For the more important individual transitions the behaviour is as follows

$$\left.\begin{array}{l} k_\nu(s\rightarrow p)\alpha v^3, \\ k_\nu(p\rightarrow s)\alpha v, \\ k_\nu(p\rightarrow d)\alpha v^5. \end{array}\right\} \qquad (11)$$

---

* A rather more quantitative discussion is given by Bates, *M.N.*, **106**, 423, 1946.

† D. R. Bates, *M.N.*, **100**, 25, 1939.

‡ B. Trumpy, *Z. Phys.*, **71**, 720, 1931.

§ It might be thought at first that at sufficiently high energies a zero would always appear. This, however, is not the case owing to the contribution C from the third major region of high density of the free wave function (neglected in the simple treatment) which is of the same sign as A and may be sufficient to compensate for the relative growth of B.

‖ D. R. Bates, *Proc. Roy. Soc.*, A, **188**, 350, 1947.

¶ D. R. Bates and H. S. W. Massey, *Proc. Roy. Soc.*, A, **177**, 281, 1941.

** *Negative Ions*, Cambridge, 1938.

In the case of resonance, however (Bates and Massey *), the situation is profoundly altered by the introduction of an additional factor of the form v^{-2} (exact) or $(a+bv)^{-2}$ (near).

At higher frequencies the cross-section (unless the transition be a resonant one from a p to an s orbital) rises to reach a maximum about where there is the closest coincidence between the wave functions—the more relatively diffuse the discrete wave function the larger the magnitude of the maximum (for a given E_0) and the less above the limit is the position at which it occurs.

The magnitude of the cross-section varies greatly from system to system but is usually between 10^{-20} cm.^{2} and 10^{-16} cm.^{2}. In general it is greatest for negative ions and least for positive ions, as would be expected.

4. *The Individual Systems.*

4.1. *Neutral Atoms.*

4.1.1. *Hydrogen.*—The problem of absorption by the hydrogen atom can be solved analytically. Even before the development of quantum mechanics Kramers† obtained his well-known formula. Later his work was confirmed and extended by, amongst others, Sugiura‡, Gaunt§, and Menzel and Pekeris.‖ The complete expression for the cross-section is very complicated, but a simplified form, which is adequate for most purposes, is

$$K_{\nu(n\Sigma l)} = \frac{32\pi^2 e^6 R}{3(3)^{\frac{1}{2}} ch^3 \nu^3 n^5} g, \qquad (12)$$

where g differs but little from unity near the spectral head but is smaller at higher energies. Actual values in a few cases are given in Table II.

TABLE II

Initial state	Cross-section at spectral head (cm.^{2})
$1s$	$6{\cdot}3 \times 10^{-18}$
$2s$	$1{\cdot}5 \times 10^{-17}$
$2p$	$1{\cdot}4 \times 10^{-17}$
$3s$	$2{\cdot}6 \times 10^{-17}$
$3p$	$2{\cdot}6 \times 10^{-17}$
$3d$	$1{\cdot}8 \times 10^{-17}$

Attempts have been made to treat other atoms as hydrogen-like with some effective nuclear charge Ze (which introduces a multiplying factor Z^4 into (12)), but these are completely unjustified (except perhaps for absorption by atoms in excited states taking Z as unity). Bates¶ pointed out that it is necessary at least to use a different value of Z for the bound and free wave functions and later** gave an approximate general formula.

4.1.2. *Helium.*—Wheeler†† and Vinti‡‡ have investigated absorption by helium. The former (whose work is the more exact) employed an eight-parameter variational function for the bound state and a simple hydrogen-like function for the

* *Phil. Trans.*, A, **239**, 269, 1943.
† *Phil. Mag.*, **46**, 837, 1923.
‡ *J. Phys. Radium*, **8**, 113, 1927.
§ *Phil. Trans.*, A, **229**, 163, 1929.
‖ *M.N.*, **96**, 77, 1935.
¶ *M.N.*, **100**, 25, 1939.
** *M.N.*, **106**, 423, 1946.
†† *Phys. Rev.*, **43**, 258, 1933.
‡‡ *Phys. Rev.*, **44**, 524, 1934.

free state. Though this last approximation might appear at first to be unduly crude, it is actually quite satisfactory. The cross-section obtained at the spectral head is $7{\cdot}4 \times 10^{-18}$ cm.^{2}. Initially it falls off more slowly than the ν^{-3} law. Here, owing to the relative compactness of the bound wave function, the cancellation is inappreciable and the results should be reliable. At high energies the accuracy is more doubtful but there is in fact fair agreement with an experimental determination by Dershem and Schein* on the absorption of the $K\alpha$ (44·6) line of carbon. Goldberg † has treated absorption by helium in the four excited states.

4.1.3. *The Alkali Metals.*—A great amount of theoretical work has been devoted to the alkali metals, partly because of their simple structure (single *s* valency electron) and partly because, owing to their low ionization potentials and ease of vaporization, they have been the main subject of laboratory measurement. For lithium the cross-section has been calculated by Hargreaves ‡ and by Trumpy §, for sodium by Trumpy §, and for potassium by Phillips || and Bates ¶. The simple self-consistent field approximation was used except for potassium, which Bates treated more elaborately with a bound wave function allowing for exchange and a *family* of free wave functions obtained by introducing a variable parameter having the form of a polarizability correction. Table III summarizes the results obtained.

TABLE III

Alkali metal	Cross-section at spectral head (cm.^{2})	Type of frequency variation
Lithium	$3{\cdot}7 \times 10^{-18}$	$\nu^{-(<3)}$
Sodium	$3{\cdot}1 \times 10^{-19}$	$\nu^{-(>3)}$
Potassium	So sensitive that no prediction is possible.	

The general trend down the group is what would have been anticipated. In all cases the potentialities of cancellation are serious—least so for lithium and most so for potassium. Actually, Bates, in his work on this last atom, abandoned any attempt at determining sufficiently accurate wave functions and concentrated on verifying that the observed minimum in the cross-section frequency curve (*cf.* Ditchburn, Tunstead and Yates **) could be explained. To account for the fact that the magnitude of the minimum is finite and not zero (as it would be on a simple approximation), allowance had to be made for non-separability. It is of interest that for the wave functions giving the experimental value of the cross-section ($1{\cdot}2 \times 10^{-20}$ cm.^{2} at the spectral head) the positive and negative contributions to the transition integral differed by only one part in fifty. The considerable influence of the polarizability correction, as shown by the following figures, is also worth noting.

Polarizability (cm.^{3})	Cross-section at spectral head (cm.^{2})
$0{\cdot}0 \times 10^{-24}$	$6{\cdot}4 \times 10^{-18}$
$0{\cdot}5 \times 10^{-24}$	$2{\cdot}8 \times 10^{-18}$
$1{\cdot}0 \times 10^{-24}$	$7{\cdot}6 \times 10^{-19}$
$1{\cdot}5 \times 10^{-24}$	$2{\cdot}2 \times 10^{-20}$
$2{\cdot}0 \times 10^{-24}$	$2{\cdot}5 \times 10^{-19}$

* *Phys. Rev.*, **37**, 1238, 1931. † *Ap. J.*, **90**, 414, 1939.

‡ *Proc. Camb. phil. Soc.*, **25**, 75, 1928.

§ *Z. Phys.*, **71**, 720, 1931; A. Damgaard has drawn my attention to a more recent calculation on sodium by M. Rudkjøbing (*Publ. Kbh. Obs.*, **18**, 1, 1941).

|| *Phys. Rev.*, **39**, 905, 1932. ¶ *Proc. Roy. Soc.*, A, **188**, 350, 1947.

** *Proc. Roy. Soc.*, A, **186**, 386, 1943.

No calculations have been carried out for either rubidium or for caesium and it is doubtful if any would be justified since both would be expected to be very sensitive to the wave functions. The existence of great cancellation is confirmed by the observed cross-sections—the values at the spectral head being only $1{\cdot}1 \times 10^{-19}$ cm.^{2} for rubidium and only between 2 and 6×10^{-19} cm.^{2} for caesium. In addition the minimum effect is observed, at least for the latter. (See review by Page *).

4.1.4. *The Alkaline—Earth Metals.*—The general approximate formula (Section 4.1.1) has been applied to beryllium (Bates †) and gives a cross-section which is $8{\cdot}2 \times 10^{-18}$ cm.^{2} at the spectral head and which falls off more rapidly than the ν^{-3} law. Calcium has been investigated in some detail by Bates and Massey.‡ The effect of exchange on both the bound and free wave functions was taken into account for the first time and was found to be important. A cross-section closely following the ν^{-3} law and of magnitude $2{\cdot}5 \times 10^{-17}$ cm.^{2} at the spectral head was obtained. In spite of the accuracy of the wave functions, the reliability of the result (though better than for potassium, the corresponding alkali metal) is not high, since the possibility that considerable cancellation might be brought about by small changes cannot be dismissed.

It is interesting to compare the construction of the low energy transition integrals of the alkali metals and the alkaline earth metals. This can most readily be done, as below, by considering the relative contributions arising from the various loops of the free wave functions (numbered from the nucleus outward):

Atom	Loop giving largest contribution	Major competing loop	Atom	Loop giving largest contribution	Major competing loop
Li	3	2	*Be*	2	3
Na	3	4	*Mg*	No calculations have been made.	
K	3	4 (almost exact cancellation)	*Ca*	3	4

The differences arise, of course, because of the relative diffuseness of the bound wave functions of the alkali metals as compared with those of the alkaline earth metals. In conjunction with the discussion in Section 3 they account, for example, for the result that the frequency variation of the cross-section for lithium is slower than that for beryllium, whereas the frequency variation of the cross-section for potassium is more rapid than that for calcium. As mentioned above, the calculations on calcium are not in themselves sufficiently trustworthy to enable the cross-section given to be accepted without reserve. Yet, in fact, they may not be greatly in error since the laboratory measurements, that show the cancellation to be almost exact for potassium, are also evidence that it cannot be high for the more compact calcium atom. Predictions for magnesium are clearly difficult to make and a detailed investigation would be of great value—though the sensitivity of the cross-section may well be prohibitively great.

* *M.N.*, **99**, 385, 1939.
† *M.N.*, **106**, 423, 1946.
‡ *Proc. Roy. Soc.*, A, **177**, 281, 1941.

It is doubtful if the heavier alkaline earth metals, strontium and barium, could be treated successfully at present.

4.1.5. *Elements with outer p^x Shells.*—Bates, Buckingham, Massey and Unwin* and Yamanouchi and Katani † calculated the cross-section for oxygen using the self-consistent field method. It was apparent from their results that the transition $p \rightarrow d$ is much more important than the transition $p \rightarrow s$. For this, fortunately, not only is there very little cancellation but also the free orbital is so dominated by the so-called centrifugal potential that it is not greatly affected by the atomic field. Bates ‡, therefore, without the labour of obtaining additional free wave functions, was able to extend the investigation to include the neighbouring elements boron, carbon, nitrogen, fluorine and neon (the bound wave functions of which are of course available). His results, which are in complete agreement with the conclusions of Section 3, are given in Table IV. They cannot be greatly in error despite the nature of the approximation.

TABLE IV §

Atom	Magnitude of cross-section at spectral head (in units 10^{-17} cm.^{2}) and form of frequency variation.
B	1·9 Medium fall-off with increase in frequency.
C	2·4 } Slow fall-off with increase in frequency.
N	2·2 }
O	$0{\cdot}4_5$ } Almost independent of frequency in range studied ($\epsilon = 0 \rightarrow 1$) (apart from discontinuities at ^2D and ^2P ionization potentials of *O* and ^1D and ^1S ionization potentials of *F*).
F	$0{\cdot}6_1$ }
Ne	$0{\cdot}5_8$ Initial slow rise with increase in frequency.

Aluminium, silicon, phosphorus, sulphur, chlorine and argon have not been investigated but they should not present undue difficulty as once again the major of the two transitions should not be unduly affected by cancellation. It is expected that the cross-sections will be more inclined to fall off at high frequencies.

The later elements of this type are less promising.

4.1.6. *The Elements with outer $d^{x'}s^{x''}$ Shells.*—The long periods contain elements such as manganese, iron and silver with outer $d^{x'}s^{x''}$ shells. Ionization at any of the many lower potentials involves either single or double transitions of the *s* electrons. Quite apart from the complexity of the systems and the probable interactions between near levels, calculations would encounter difficulties as with the heavier alkali and alkaline earth metals and would certainly not be reliable. All that can be stated at present is that in many instances the cross-sections must be low.

* *Proc. Roy. Soc.*, A, **170**, 322, 1939.

† *Proc. phys. math. Soc. Japan*, **22**, 60, 1940.

‡ *M.N.*, **100**, 25, 1939.

§ In the original paper the main interest was felt to lie in the trend of the transition integrals from boron to neon and in consequence only the cross-sections per $2p$ electrons (assumed to be identical) were given. ϵ is in Hartree units (13·53 e. volts).

4.2. *Positive Ions.*—There has been scant theoretical study of the positive ions. Formula (12) (with the Z^4 multiplying factor) applies of course to hydrogen-like systems, and for He^+ gives a cross-section of $1{\cdot}6 \times 10^{-18}$ cm.^{2} at the spectral head. A preliminary estimate can also be made of the cross-section of Li^+, as for it the effective nuclear charge on the bound and free orbitals is not very different and as the cancellation is not very serious. The magnitude at the spectral head appears to be about 2×10^{-18} cm.^{2}, but confirmation by a proper calculation is necessary.

Only a single case, the alkali-metal-like ion Ca^+, has been investigated in detail. Using bound and free wave functions corrected for exchange (which again is of importance), Bates and Massey* found a cross-section rising slowly with frequency from a value of $2{\cdot}4 \times 10^{-19}$ cm.^{2}. Unfortunately little reliance can be placed on the result owing to the great sensitivity to details of the wave functions.

Bates † has employed the general approximate formula (Section 4.1.1) in the study of C^+, N^+, O^+, F^+, Ne^+ and Na^+, all of which have of course an outer $2p^x$ shell. There is little cancellation and the accuracy of the magnitudes and frequency variations of the cross-sections obtained should be fair. The results are summarized in Table V—comparison with Table IV is instructive.

TABLE V

Ion	Magnitude of cross-section at spectral head (in units 10^{-18} cm.^{2}) and form of frequency variation).
C^+	3·7
N^+	6·4 — Slow fall-off with increase in frequency.
O^+	8·1
F^+	2·5
Ne^+	4·5 — Almost independent of frequency in range studied (ϵ=0→1) (apart from discontinuities at ^2D and ^2P ionization potentials of F^+, and ^1D and ^1S ionization potentials of Ne^+).
Na^+	7·1

In general it would appear that as for neutral atoms the immediate useful scope of theory is rather limited by the cancellation phenomenon—and also of course by the complexity of the heavier systems. However, some further cases should be treated and among the more suitable may be mentioned Be^+, B^+, Mg^+, Al^+ (outer s^x shells) Si^+, P^+, S^+, Cl^+, A^+ and K^+ (outer p^x shells).

4.3. *Negative Ions.*

4.3.1. On account of its astrophysical importance a great amount of work has been directed to H^-. The first attempt at precise calculation was made by Massey and Bates.‡ For the bound state (which has configuration (s^2)) they accepted the existing three parameter variational function and for the free state they showed that the interaction between the ejected electron and residual hydrogen atom could be neglected—the effects of the normal field, of exchange, and of polarization, were all examined and were proved to be small. The computation was repeated by Williamson § using an improved bound wave function involving six parameters. While this was an advance on the work of Massey and Bates,

* *Proc. Roy. Soc.*, A, **177, 281, 1941.**
† *M.N.*, **106,** 423, 1946.
‡ *Ap. J.*, **91,** 202, 1940.
§ *Ap. J.*, **99,** 438, 1942.

Chandrasekhar and Krogdahl * found reason to question the accuracy from the application of sum rule tests.† Henrich ‡ introduced eleven parameters but even his results were doubtful. In a review of the position Chandrasekhar § emphasized that the essential difficulty is that the variational method yields a good approximation to the wave function only in the region giving the main contribution to the energy integral whereas other regions are important in the transition integral. Later || he pointed out that while of course the expressions for the cross-section corresponding to the dipole length, velocity and acceleration are formally identical, in practice they are not of the same merit owing to their dependence on different parts of the wave functions. For the system concerned the matrix elements are respectively,

$$\left|\int_\tau \Psi^{*i}(z_1+z_2)\Psi^f d\tau\right|, \tag{13}$$

$$\left|\frac{1}{h\nu}\int_\tau \Psi^{*i}\left(\frac{\partial\Psi^f}{\partial z_1}+\frac{\Psi\partial^f}{\partial z_2}\right)d\tau\right| = \left|\frac{1}{h\nu}\int_\tau \left(\frac{\partial\Psi^{*i}}{\partial z_1}+\frac{\partial\Psi^{*i}}{\partial z_2}\right)\Psi^f d\tau\right|, \tag{14}$$

and

$$\left|\frac{1}{(h\nu)^2}\int_\tau \Psi^{*i}\left(\frac{z_1}{r_1^3}+\frac{z_2}{r_2^3}\right)\Psi^f d\tau\right|. \tag{15}$$

Chandrasekhar found that loss of accuracy occurs with (13) due to the emphasis on large radial distances and with (15) due to the emphasis on small radial distances. He therefore adopted (14) as the most suitable form and again repeated the calculation. The final results ¶ show the cross-section to rise from zero to a maximum of $4{\cdot}4\times10^{-17}$ cm.^{2} at $\lambda 8500$.

In certain other cases the dipole velocity might also profitably be employed. A possible disadvantage, however, is that it involves the first derivative of one or other of the wave functions. The degree of cancellation must also be taken into account.

Potentially there are two main families of negative ions having outer s^x shells—those of the inert gases and those of the alkali metals. The former, however, probably do not exist as stable entities and need not be considered here. A detailed calculation of the cross-section of one of the latter, Na^-, has recently been carried out by Huang. ** The bound wave function included the effect of exchange and is the best approximation at present available. Despite this, it is unlikely to be accurate. In the free wave function it was found, as with H^-, that the ejected electron can be represented adequately by a plane wave. Owing to the uncertainty of the detachment energy Huang employed two values 0·0 and 0·5 e. volts, which should be lower and upper limits respectively. Though possibly considerably in error the results obtained are of great interest as being typical of those of the larger negative ions in which $s\rightarrow p$ transitions are involved. In accordance with expectation (Section 3) the cross-section increases from zero

* *Ap. J.*, **98**, 205, 1943.

† The sum rate tests are not generally very useful owing to the excited states but for H^- these are assumed not to exist.

‡ *Ap. J.*, **99**, 59, 1944.

§ *Rev. Mod. Phys.*, **16**, 301, 1944.

|| *Ap. J.*, **102**, 223, 1945.

¶ *Ap. J.*, **102**, 395, 1945.

** *Ap. J.*, **101**, 196, 1945.

to reach a huge maximum slightly above the spectral head and then decreases rapidly. Huang actually does not give the magnitude and position of the maximum but the computation can readily be done to yield the following:—

Detachment energy assumed	Maximum cross-section (cm.^{2})
0·0 e. volts	$0{\cdot}9 \times 10^{-16}$ at λ 25,000
0·5 e. volts	$2{\cdot}0 \times 10^{-16}$ at λ 15,000

Calculations on the negative ions of the other alkali metals are scarcely justified since, except for Li^-, the detachment energy is unknown, and since in all cases the bound wave function is poorly determined.

Isolated instances of anomalously placed negative ions with outer *s* shells such as Be^- and Cu^- may occur but the information on them is slight and detailed theoretical investigation impractical.

4.3.2. Bates and Massey * and Bates † have discussed absorption by O^- ($2p^5\,^2P$). They used a bound wave function allowing for exchange. Of the relevant free wave functions, those relating to the *d*-orbital could readily be computed owing to the dominance of the centrifugal effect, but those relating to the *s*-orbital (where of course this is absent) required very careful investigation—especially in view of the possibility of resonance ‡ (Section 3). Besides exchange corrections, a polarizability parameter was introduced, and calculations were carried out for a wide range of values of it. These showed that resonance only occurs if the polarizability is several times larger than would have been anticipated. While this contingency is thought to be improbable §, it cannot be excluded, and in giving their results Bates and Massey made allowance for it. The cross-section curves are by no means simple but the main features are briefly as follows:—

(i) Non-resonance. Cross-section increases from zero at the spectral head to a broad maximum of about 2×10^{-17} cm.^{2} at rather below λ 1000. Discontinuities of slope occur at each of the detachment limits.

(ii) Resonance. Similar to (i) except just above detachment limits where the cross-section tends to become infinite.

Unfortunately the results depend greatly on the assumed bound wave function at large radial distances where it is far from reliable.

Cl^- ($2p^6\,^1S$) also has been investigated (Massey and Smith‖), though not in the same detail as O^-. Using wave functions derived from the self-consistent field approximation, a cross-section curve of the usual form with a maximum of magnitude $2{\cdot}1 \times 10^{-17}$ cm.^{2} occurring at λ 1200 was obtained.

Calculations on other negative ions with outer *p* shells have not been carried out and in view of the inaccuracy of the results that can be achieved at present they are barely warranted now that the main possibilities are established.

* *Phil. Trans.*, A, **239**, 269, 1943.

† *M.N.*, **106**, 128, 1946.

‡ Resonance was not mentioned in the previous section as its occurrence can probably be discounted except where the free azimuthal quantum number is zero.

§ There is evidence in favour of the existence of excited states of O^- (W. W. Lozier, *Phys. Rev.*, **46**, 268, 1934, and H. T. Hagstrum and J. T. Tate, *Phys. Rev.*, **59**, 354, 1941) but if their stability is appreciable they may arise from the augmenting of the normal field by distortion rather than by inordinate polarization of the core.

‖ H. S. W. Massey and R. A. Smith, *Proc. Roy. Soc.*, A, **155**, 472, 1936.

5. *Summary of Present Position.*—The following lists show the various systems of the Periodic Table and give:—

(1) Whether or not calculations have been performed.

(2) (*a*) If calculations have been performed the probable reliability and the possibility of improvement.

(*b*) If calculations have not been performed the estimated reliability of any likely to be carried out with present methods.

The grouping, though broad and rather arbitrary, gives a useful outline of the situation.

Neutral Atoms	Calculations performed	Comment
H	Yes	Exact.
He	Yes	Fair. Could be improved.
Alkali metals: *Li, Na*	Yes	Fair. Could be improved.
Alkali metals: *K*	Yes	Poor. Improvement very difficult.
Alkali metals: *Rb, Cs*	No	Reliability likely to be poor.
Alkaline earth metals: *Be*	Yes	Fair could be improved.
Alkaline earth metals: *Mg*	No	Fair reliability may be possible (but occurence of high sensitivity cannot be dismissed).
Alkaline earth metals: *Ca*	Yes	Fair. Improvement difficult.
Alkaline earth metals: *Sr, Ba*	No	Reliability likely to be poor.
Systems with outer p^x shells: *B, C, N, O, F, Ne*	Yes	Fair. Could be improved.
Systems with outer p^x shells: *Al, Si, P, S, Cl, A*	No	Fair reliability should be possible.
Remaining neutral atoms	No	Not at all promising.
Positive Ions		
He^+	Yes	Exact.
Li^+	Yes	Poor. Could be improved.
Alkali metal-like ions: Be^+, Mg^+	No	Fair reliability may be possible.
Alkali metal-like ions: Ca^+	Yes	Poor. Improvement very difficult.
Alkali metal-like ions: Sr^+, Ba^+	No	Reliability likely to be poor.
Alkaline earth metal-like ions: B^+, Al^+	No	Fair reliability may be possible.
Alkaline earth metal-like ions: Ga^+, In^+,	No	Reliability likely to be poor.
Systems with outer p^x shells: C^+, N^+, O^+, F^+, Ne^+, Na^+	Yes	Fair. Could be improved.
Systems with outer p^x shells: Si^+, P^+, S^+, Cl^+, A^+, K^+	No	Fair reliability should be possible.
Remaining positive ions	No	Not at all promising.

Negative Ions

The only negative ion for which the calculations can be regarded as at all accurate is H^-. While Na^-, O^- and Cl^- have been studied the results obtained (though giving valuable information on representative types) may be considerably in error. None of the remaining negative ions show any more promise.

6. *Conclusion.*—A sufficient number of individual systems have been investigated to give the general trend of absorption cross-sections. The main future development of the theory probably lies in attempting to achieve increased accuracy. This requires a careful study of the importance of the corrections to the wave functions to allow for the effects of exchange, polarizability of the core and non-separability. Further the most suitable form of quantal expression for the cross-section must be investigated so that in any individual case the influence of the expected deficiences in the wave functions is minimized. The great computational labour involved is likely severely to limit any programme.

Finally I would like to thank Professor Massey for many helpful discussions.

Department of Mathematics,
University College,
*London, W.C.*1:
1946 *May* 23.

ALFVÉN'S THEORY OF SUNSPOTS

T. G. Cowling

(Received 1946 September 25)

Summary

Alfvén's theory of sunspots is reviewed critically. The theory is shown to be, at a number of points, internally inconsistent, or inconsistent with observation. It is concluded that, while certain details of the theory may have permanent value, the theory as a whole is not acceptable.

---

1. In a series of papers published recently, H. Alfvén * has advanced a new theory of sunspots. This theory he uses to explain many properties of spots, such as their tendency to occur in pairs of opposite magnetic polarity, the duration of the solar cycle, the reversal of polarities in each new cycle, the progression of spots towards the equator during a cycle, etc. His work has been amplified and developed by C. Walén.†

The theory includes two hypotheses.

Hypothesis A.—The darkness of a spot is due to the mechanical effect of its magnetic field.

Hypothesis B.—The field is generated from the Sun's general magnetic field by reason of wave disturbances transmitted along the lines of force.

The first of these hypotheses does not necessarily involve the second, but the second without the first would imply that magnetic fields would not necessarily be associated with sunspots. In what follows, these two hypotheses will be examined critically.

2. For a mathematical discussion of the theory, we must refer to Walén's paper; an adequate understanding of the theory can, however, be reached without detailed mathematics by appealing to two known physical laws.

Law I.—The mechanical forces acting on non-magnetic material as a result of a magnetic field H is the same as if the lines of force were subject to a tension $H^2/8\pi$ per unit area of a normal cross-section, and there were a pressure of the same amount normal to the lines of force.

Law II.—Lines of force are "frozen" into highly conducting material; that is, the material can flow freely along the lines of force, but when it flows normal to the lines of force it drags them with it.

Law II is not universally valid; in the presence of a steady transverse electric field E (in E.M.U.) of electrostatic origin, a transverse velocity E/H of the material relative to the lines of force is possible. Such electric fields appear, however, to be unimportant in Alfvén's theory. The electrical conductivity of solar material is so large that the lines of force can certainly be taken to be "frozen" into the materials.

3. Alfvén originally based his Hypothesis A (that the magnetic field causes the

* *Ark. Mat. Astr. Fys.*, **29**A, Nos. 11 and 12, 1943, and **29**B, No. 2, 1942; *M.N.*, **105**, 3 and 382, 1945.

† *Ark. Mat. Astr. Fys.*, **30**A, No. 15, 1944.

coolness of a spot) on a consideration of the state of mechanical equilibrium in the presence of an intense magnetic field. Taking the field to be vertical, he found the equilibrium state to be such that both the density, and the sum of the gas pressure and the lateral magnetic pressure $H^2/8\pi$, are uniform over any horizontal plane; the gas pressure by itself, and also the temperature, are lower in the region of intense magnetic field than outside. Quantitative relations are somewhat different for other fields (the actual field of a sunspot seems to have been supposed by Alfven and Walén to resemble that shown in Fig. 1); but the general result is unaltered, that for equilibrium the temperature must be low in regions of intense magnetic field.

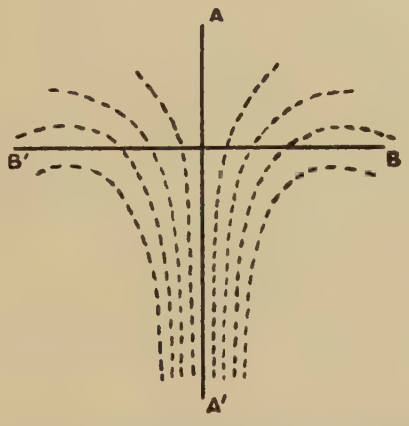


FIG. 1.—*Lines of force of the assumed sunspot magnetic field, in a vertical plane through the (vertical) axis of symmetry of the spot. The* BB′ *line represents the Sun's surface.*

Alfvén did not suggest how the equilibrium state is either attained or maintained against the heating effect of radiation. One possible mechanism for attaining it is as follows. Suppose the lines of force to be initially in a tight bundle; the lateral magnetic pressure $H^2/8\pi$ causes the bundle to expand, and, since the lines of force are "frozen" into the material, a partial vacuum is produced inside the bundle. This may suck up material from great depths along the lines of force, and there may be further expansion; finally, as a result of adiabatic cooling, the equilibrium state is reached, with the temperature lower in the bundle than outside. A visible spot is then formed.

The spot may remain cool during its lifetime either because the process of its formation creates a large reservoir of cool material which is only slowly warmed up by radiation, or because material is continuously pumped up from below during the spot's lifetime, and cools adiabatically as it rises. Walén examined these alternatives of "reservoir" and "pumping" maintenance. He showed that a reservoir of cool material remains cool for some months against the warming influence of radiation at a depth where the pressure is 10^9 dynes/cm.^{2}, but only for a second or two at the surface. Of course, the existence of a reservoir of cool material at great depths implies some reduction in the radiation reaching the surface; but the coolness at great depths implied by Alfvén's theory means a reduction of temperature by only one or two per cent. there, and cannot by itself very seriously reduce the radiation reaching the surface unless the opacity varies abnormally as a

result of a small change in temperature. Walén concluded that the surface coolness is not maintained by a reservoir effect, but by the adiabatic cooling of rising gas.

Walén calculated the velocity of the rising gas by balancing the change in pressure due to radiative warming against the dynamical pressure ($\frac{1}{2}\rho v^2$). This is extremely unsatisfactory; the dynamical pressure is one of the least important of the effects of the motion. The most important effects of the motion are thermal, and the velocity of rising should actually be calculated by balancing heat lost by adiabatic cooling against heat gained by radiation. So far as mechanical effects are concerned, the change in pressure due to adiabatic cooling is far more important than the dynamical pressure. Moreover, it is not clear that Walén's calculation pays due regard to Law II of Section 2, which implies that, at any point of the surface where the magnetic field is not directed along the surface, and in particular near the centre of the visible spot, no motion can exist save such as results from the field expanding and carrying material with it. We therefore consider the question *de novo*.

4. We assume that a field like that of Fig. 1 expands near the Sun's surface, and that material is pumped up from great depths to fill the partial vacuum thus created. We consider in what circumstances such pumping can maintain the observed central coolness of the visible spot.

Ignoring viscous stresses, the heat equation for a moving gas is

$$\frac{\partial E}{\partial t} + \operatorname{div}(E\boldsymbol{v}) - \frac{p}{\rho}\left(\frac{\partial \rho}{\partial t} + \boldsymbol{v} \,.\, \operatorname{grad}\rho\right) = -\operatorname{div}\boldsymbol{F}, \tag{1}$$

where p, ρ, $\boldsymbol{v}$ and E are the pressure, density, velocity and heat energy per unit volume of the gas, and $\boldsymbol{F}$ is the radiative flux of energy. The equation of continuity is

$$\frac{\partial \rho}{\partial t} + \operatorname{div}(\rho\boldsymbol{v}) = 0. \tag{2}$$

If γ is the ratio of specific heats, we may put

$$E = p/(\gamma - 1). \tag{3}$$

Then (1) becomes

$$\frac{1}{\gamma - 1}\left\{\frac{\partial p}{\partial t} + \boldsymbol{v} \,.\, \operatorname{grad} p - \frac{\gamma p}{\rho}\left(\frac{\partial \rho}{\partial t} + \boldsymbol{v} \,.\, \operatorname{grad}\rho\right)\right\} = -\operatorname{div}\boldsymbol{F}. \tag{4}$$

In this equation we approximate by omitting the $\partial/\partial t$ terms; this is equivalent to ignoring any reservoir effect. We also ignore the horizontal inflow of radiation, and the thermal effects of horizontal expansion. Then

$$\frac{w}{(\gamma - 1)}\left\{\frac{\partial p}{\partial \zeta} - \frac{\gamma p}{\rho}\frac{\partial \rho}{\partial \zeta}\right\} = -\frac{\partial F}{\partial \zeta},$$

where w is the vertical velocity, ζ the height above the base of the spot—a layer where the coolness of the spot can be supposed to start. Integrating through the height Z between the base of the spot and the Sun's surface,

$$\Delta F = \int_0^Z \frac{w}{\gamma - 1}\left\{\frac{\partial p}{\partial \zeta} - \frac{\gamma p}{\rho}\frac{\partial \rho}{\partial \zeta}\right\} d\zeta, \tag{5}$$

where ΔF is the difference between the radiative flux at great depths and that at the surface, i.e. between the surface flux outside the spot and that inside.

To estimate w, suppose first that the magnetic field is vertical. Then, as the lines of force expand in a cylinder of radius r round the (vertical) axis of the spot, material from below the height ζ is sucked up to replenish the cylinder above this height. Thus at the height ζ

$$\pi r^2 \rho w = \int_\zeta^Z \rho \dot{r} \,.\, 2\pi r \, d\zeta$$

$$= 2\pi r \dot{r} \int_\zeta^Z \rho \, d\zeta$$

$$= 2\pi r \dot{r} p/g,$$

where g is the acceleration of gravity; or

$$w = 2 \frac{\dot{r}}{r} \frac{p}{\rho g}. \tag{6}$$

During the middle part of the lifetime of large spots, the central field certainly does not alter by 10 per cent. in ten days.* Hence, since the total magnetic induction along our cylinder, πr^2 times the field, is constant as the field expands, $2\dot{r}/r$ must be less than 10^{-7} sec.$^{-1}$, and

$$w < 10^{-7} p/\rho g.$$

The actual magnetic field is not vertical, but resembles that of Fig. 1. Thus rising material has to replenish a greater volume near the surface; on the other hand, the field expands more slowly at great depths than near the surface. In consequence, w may equal, or even exceed, $10^{-7} p/\rho g$ near the surface; at great depths it is smaller.

In (5), we put $w = 10^{-7} p/\rho g$; this is almost certainly not an underestimate when applied to the whole range of integration. Then this equation becomes

$$\Delta F = \frac{10^{-7}}{(\gamma - 1) g} \int_0^Z \frac{p}{\rho} \left(\frac{\partial p}{\partial \zeta} - \frac{\gamma p}{\rho} \frac{\partial \rho}{\partial \zeta} \right) d\zeta. \tag{7}$$

Suppose that $p \propto \rho^n$ in the atmosphere; then

$$\Delta F = - \frac{10^{-7} (\gamma - n)}{(2n - 1)(\gamma - 1) g} \int_0^Z \frac{\partial}{\partial \zeta} \left(\frac{p^2}{\rho} \right) d\zeta$$

$$= \frac{10^{-7} (\gamma - n)}{(2n - 1)(\gamma - 1) g} \left(p \,.\, \frac{R}{\mu} T \right)_{\zeta = 0},$$

where R is the gas constant, μ the mean molecular weight of the material. Taking $\Delta F = 3 \times 10^{10}$ erg./cm.^{2} sec., $\mu = 1$, $n = \frac{4}{3}$, $\gamma = \frac{5}{3}$, this gives

$$(pT)_{\zeta = 0} = 3{\cdot}3 \times 10^{14} \text{ deg. dynes/cm.}^2.$$

The separate values of p and T are hard to estimate from this; while it is not unreasonable to assume $p \propto T^4$ in the Sun's subsurface layers, recent estimates of the photospheric pressure have varied widely. We adopt as reasonable the values $T = 100{,}000°$, $p = 3{\cdot}3 \times 10^9$ dynes/cm.^{2} for the temperature and pressure at the base of the spot. Because of the numerous approximations made in arriving at these values, and notably the neglect of the horizontal flow of radiation into the spot, these values may be underestimates, but are not likely to be overestimates. Since, with $p \propto T^4$ and $\mu = 1$, the vertical temperature gradient is about 8·3°/km.,

* *M.N.*, **106**, 218, 1946.

they correspond to the base of the spot being rather more than 11,000 km. below the Sun's surface.

This estimated depth, though far less than the diameter of a large spot, is none the less too large to be reasonable. The form of (7) implies that the major part of the reduction in F takes place near the base of a spot ; through most of the 11,000 km. between the base of the spot and the Sun's surface, the temperature gradient is nearly radiative, corresponding to a reduced F. The lateral inflow of radiation in this 11,000 km. must be important; it must produce a "fuzziness" of the edge of the umbra, which should fade off slowly into the penumbra in a belt some thousands of km. broad. In fact, however, the edge of the umbra is relatively sharp, suggesting that the origin of the darkness of the spot is only one or two thousand km. below the surface.

Again, if material between the base of the spot and the surface is, for most of its height, in radiative equilibrium with a reduced F, its pressure must differ considerably from that of surrounding material at the same level. The maximum pressure-difference between spot and surrounding material must be at least one or two per cent. of the pressure at the base—say, it must be of order 10^8 dynes/cm.^{2}. To balance this pressure-difference, there is only the lateral magnetic pressure $H^2/8\pi$, and, for this to equal 10^8 dynes/cm.^{2}, H would have to be 50,000 gauss. But, on the magnetic theory of sunspot maintenance, the cross-section of the field at the base of a spot must equal the visible spot cross-section; hence the field at great depths cannot have a much smaller cross-section than the surface field, and its magnitude cannot exceed two or three times that of the surface field.

These difficulties have arisen because of Law II of paragraph 2, which implies that vertical velocities below a sunspot must be very small. Theory suggests no reason why this law should cease to hold near the surface; but, even if it did, the difficulties would persist in an altered form. A large part of solar material is hydrogen, which is neutral at the surface, but ionized one or two thousand km. lower. Its energy of ionization was ignored when γ was taken to be $\frac{5}{3}$ in (7); this was permissible because, of the total mass of rising gas, only a small part rises through the zone in which hydrogen ceases to be ionized. But if Law II were not valid at the surface, and vertical velocities below the spot were much greater than we have supposed, the motion would have to approximate rather to a circulation such that the total mass of rising gas in the spot is much the same at all depths. The ionization energy surrendered as the gas nears the surface is more than enough to compensate for the whole of the adiabatic cooling as the gas rises from a depth where the temperature is 100,000°; thus, to produce a dark spot, gas must rise from yet greater depths. This only aggravates our difficulties.

5. It is also difficult to explain the observed speeds of Evershed flow (nearly horizontal flow out of the spot umbra), if this is due solely to mechanical effects of the magnetic field. The flow takes place, on this theory, because the magnetic pressure pushes the lines of force outward, and these drag the material with them. If, owing to the finite conductivity of the material, the lines of force are not firmly "frozen" into it, the material tends to lag behind the lines of force, not to travel faster. Thus the maximum outward speed is determined by the rate of expansion of the field. The field can hardly expand more than 10,000 km. in ten days, or 0·01 km./sec., as against Evershed speeds of order 1 km./sec.

The gap between these two speeds can, though with difficulty, be bridged by making two assumptions. (i) Solar material moves freely along the lines of force;

only its transverse motion is determined by the expansion of the field. Thus if the field is inclined to the Sun's surface at only four or five degrees, the motion of the material (along the surface) may be ten times as great as that of the field (transverse to itself). (ii) The field at the centre of a spot does not vary greatly during the middle part of the spot's lifetime; the expansion of the field is largely a separation of the wing lines of force.* If the Evershed velocities arise in a region where the field is only one-tenth that at the spot centre, the rate of expansion of the field here may be greatly increased without an undue diminution of the central field resulting.

Both these assumptions must be made in order to bridge the gap. Observations to check their validity are greatly to be desired. In measuring the inclination of the field to the Sun's surface, it must be remembered that the surface is not necessarily horizontal near a spot; observations need to check whether the inclination of the field to the Evershed velocity is small.

Of course, if the Evershed velocity is not due to motion of the lines of force, the difficulties about its magnitude are considerably lessened: the magnetic field must still be nearly parallel to the direction of flow, but it need not be small in the region of flow. But Alfvén's theory, it must be stressed, demands that the flow arises from mechanical effects of the field, and so from the motion of the lines of force.

6. We turn to consider Alfvén's Hypothesis B (Paragraph 1), that the sunspot magnetic field is generated from the Sun's general field by reason of "magneto-hydrodynamic" (m.-h.) waves transmitted along the lines of force of that field. In his theory of m.-h. waves, Alfvén assumes (electrically) perfectly conducting material. I find his assumption to be not unreasonable in the Sun. For, using the conductivities of solar material given in an earlier paper †, electrical resistance can be shown to lead to little loss of wave energy in a disturbance whose dimensions are larger than 1000 km., as the disturbance travels from near the centre of the Sun to the surface.

In most wave motions, the wave disturbance is small. In m.-h. waves, on the other hand, the disturbance field (that which manifests itself in the spot field) is large compared with the undisturbed field, so that the wave is one of large amplitude, like a shock wave. This point was examined in detail by Walén, who showed that the largeness of the disturbance produces little dissipation of energy, or change of wave form, in any region where the undisturbed field and the density vary slowly from point to point.

Hypothesis B, as advanced by Alfvén, was incomplete; it did not discuss the form of the wave disturbance, nor its behaviour near the Sun's surface. Walén did discuss these points, and we shall therefore confine attention to the theory as developed by him. But many of the considerations which will later be advanced would apply with equal force to any version of the theory.

Walén supposed that instability near the centre of the Sun leads to a "rotating cartwheel" vortex disturbance. This breaks up into two disturbances, propagated in opposite directions along the lines of force of the Sun's general field. The types of motion $\boldsymbol{v}$ and field $\boldsymbol{H}$ in one of these disturbances are indicated in Fig. 2. The motion $\boldsymbol{v}$ is confined to a tore, or anchor-ring, and is a motion purely round the axis of the tore; in Fig. 2 the axis of the tore is taken parallel to the undisturbed

* *M.N.*, **106**, 218, 1946.
† *M.N.*, **105**, 166, 1945.

field, but this is not essential. As the tore advances through the material, lines of force are carried round the tore by moving matter; a typical line of force is that which is along CD before the tore passes, and is carried round to the line AB during its passage. In the tore, the field $\boldsymbol{H}$ is much more intense than the undisturbed field, and is nearly parallel to $\boldsymbol{v}$.

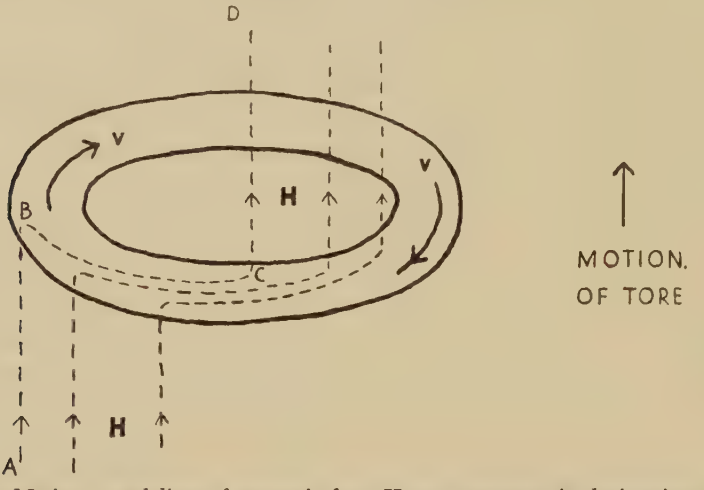


FIG. 2.—*Motion* $\boldsymbol{v}$ *and lines of magnetic force* $\boldsymbol{H}$ *as a tore moves in the interior of the Sun. Lines of force are shown as broken lines; matter circulates round the tore. The tore does not carry matter with it, but moves through the matter.*

The transverse magnetic pressure $H^2/8\pi$ and the equal magnetic tension along the lines of force are large inside the tore. The transverse magnetic pressure is balanced by a decrease in gas pressure, the effect of this decrease along the lines of force being equivalent to increasing the magnetic tension to $H^2/4\pi$. The propagation of the tore is due to this tension. At a point like C on a line of force, the tension jerks the material in front of the advancing tore into motion: at a point like B, the tension stops the movement of material at the back of the tore. Inside the tore, the tension balances centrifugal force.

The speed of advance of the tore is $H_0/\sqrt{(4\pi\rho)}$, where H_0 is the undisturbed field; this speed increases rapidly as the surface is approached. The plane of the tore is in general not horizontal; the tore's upper part moves faster than the rest near the surface, producing a distortion of the tore. As parts of the tore successively reach the surface, they are reflected; during the reflection, the motion and field are shown in Fig. 3. At any instant, the tore meets the surface in two areas over which the field has opposite senses; these, according to Hypothesis A, are the two spots of a bipolar group. The motion in the tore is in the same sense round the tore at all points; the field in the tore is in opposite senses round the tore on opposite sides of the spots.

7. A number of objections can be urged against this hypothesis.

(i) To explain the solar cycle, a more or less regular alternation of stability and instability in the Sun's core must be postulated. I have criticized the theory of this alternation elsewhere.*

(ii) The alternation in magnetic polarity of spots from cycle to cycle is explained

* *M.N.*, **105**, 166, 1945 (Appendix).

by a periodic change in the Sun's internal angular velocity. The arguments adduced in favour of such a periodic change are obscure and unconvincing.

(iii) Walén is obliged to assume an initial "rotating cartwheel" vortex in which the rotation is confined to the rim, and absent near the axis. Vortices actually observed in disturbed fluid, on the other hand, have their greatest angular velocity near the axis.

(iv) Walén assumed that a tore, when formed near the Sun's centre, has its axis parallel to the Sun's axis. Alfvén * pointed out that, if this is so, the spots of a

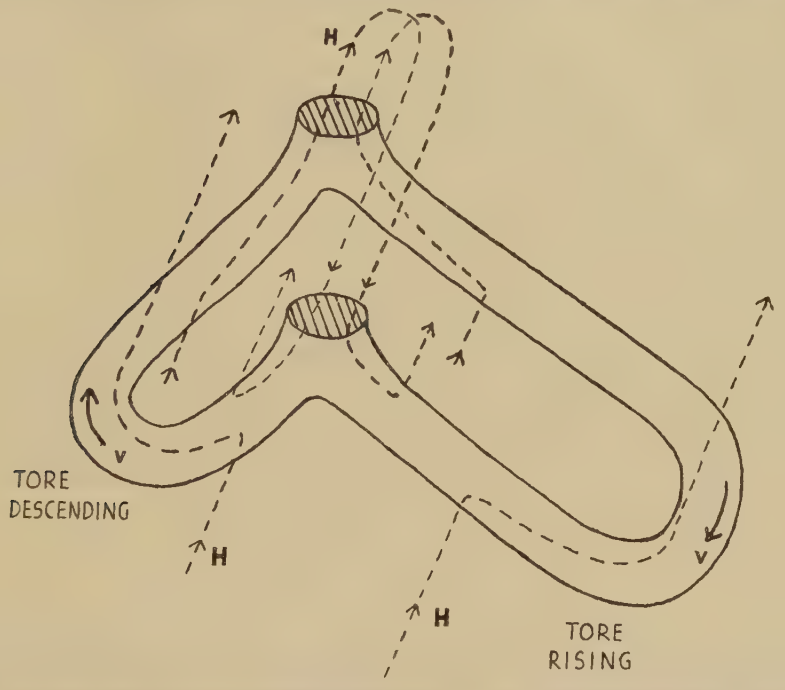


FIG. 3.—*Motion and lines of force during the reflection of a tore at the Sun's surface. The shaded areas are the sunspots in which the tore meets the surface.*

large bipolar group should move some degrees equatorward during their lifetime, which is contrary to observation. He suggested instead that the plane of the tore, at its formation, must be parallel to the Sun's general magnetic field, and supported his suggestion by showing that, when this is so, the rate at which the tore increases in energy during its formation is greatly enhanced. There are serious difficulties in Alfvén's discussion; he assumes that the tore is propagated during its formation in exactly the same way as if it were already completely formed, and wholly neglects the forces increasing the tore's energy in considering its motion. But, even accepting his argument, agreement with observation is not obtained. The forces which increase the energy of the tore are vertical forces, and so must lead to motion in a vertical plane; that is, the plane of a tore should be a vertical plane parallel to the Sun's general magnetic field, i. e. a meridian plane. This remains true during the tore's passage to the Sun's surface; thus the tore should lead to two spots in the same meridian plane, and the spots of a bipolar group should have the same longitude, but different latitudes.

* *M.N.*, **105**, 382, 1946.

(v) Walén supposed a disturbance near the Sun's centre to generate two equal tores, which reach the surface in opposite hemispheres. Thus groups of spots should occur in roughly equal pairs, one in each hemisphere. The groups of a pair need not appear at exactly the same time, but the total numbers and areas of spots occurring in the two hemispheres in one cycle should be roughly equal. This is contrary to observation. In the latter part of the seventeenth century and the beginning of the eighteenth, the Sun's northern hemisphere failed to produce a single recorded spot; the southern did produce spots, though few. More recent sunspot cycles (e.g. that with maximum in about the year 1883) also showed a marked difference between the behaviours of the two hemispheres.*

This criticism is, however, weakened if a tore is not generated suddenly, but steadily increases in energy, as Alfvén suggested (see (iv), above).

(vi) In order to produce visible spots when it reaches the surface, the tore must not be in a horizontal plane. But if it is not in a horizontal plane, its passage in general displaces material vertically. At any level where the Sun is stable, energy is required to produce such a vertical displacement, and the energy required is not negligible.

Because tubes of magnetic force are constant in strength along their lengths, the ratio of the field in the tore in Fig. 2 to the field outside is the ratio of the distance a line of force travels in the tore to the width of the tore's cross-section. The surface radius of a large spot is, say, 10,000 km., and its central field is 2,500 gauss; if the cross-sectional radius of the tore is $10{,}000f^{-1}$ km., the field in it is of order $2500f^2$ gauss. The field outside the tore is, in the outer layers of the Sun, of order 100 gauss. Hence, in the outer layers, the distance a line of force travels in the tore is of order

$$\frac{2{,}500f^2}{100} \times \frac{10{,}000}{f} = 250{,}000f \text{ km.},$$

and this is the distance which material moves round the tore as it passes. The vertical part of this distance must be large if the tore is not horizontal, and so an appreciable amount of work must be done at the expense of the energy of the m.-h. wave.

The rate of loss of m.-h. energy can be calculated from results given by Alfvén.† He finds that, as the tore passes by, the gravitational energy gained by unit volume of the material through which the tore passes is $\Theta(\zeta_1-\zeta_0)^2$, where $\zeta_1-\zeta_0$ is the distance which the material rises, and

$$\Theta = \tfrac{1}{2}g\left(-\frac{\partial\rho}{\partial\zeta}+\frac{\rho}{\gamma p}\frac{\partial p}{\partial\zeta}\right) = -\tfrac{1}{2}g\frac{\rho}{p}\frac{\partial p}{\partial\zeta}\left(\frac{1}{n}-\frac{1}{\gamma}\right), \tag{8}$$

using the same notation as in Paragraph 3. Thus when the tore moves a distance equal to its cross-sectional width a, it loses energy to the surrounding material of amount $V\Theta\overline{(\zeta_1-\zeta_0)^2}$, where V is its volume, and bars indicate mean values. In travelling a distance equal to the "scale-height" $p(-\partial p/\partial\zeta)^{-1}$ of the pressure variations, the energy lost is

$$V\Theta\overline{(\zeta_1-\zeta_0)^2}\,p(-a\partial p/\partial\zeta)^{-1} = \tfrac{1}{2}Vg\rho(\zeta_1-\zeta_0)^2\left(\frac{1}{n}-\frac{1}{\gamma}\right)\Big/a.$$

* See Maunder, *M.N.*, **82**, 534, 1922.

† *M.N.*, **105**, 382, 1945.

The ratio of this to the whole energy of the tore is *

$$\tfrac{1}{2}g(\zeta_1-\zeta_0)^2\left(\frac{1}{n}-\frac{1}{\gamma}\right)\Big/av^2,$$

where v is the material velocity. Taking $a=10{,}000$ km., $v=5\times10^4$ cm./sec., $g=4\times10^4$ cm./sec.^{2}, $n=\frac{4}{3}$, $\gamma=\frac{5}{3}$ (values not unreasonable 100,000 km. below the surface) this is $1{\cdot}2\times10^{-15}(\zeta_1-\zeta_0)^2$. Thus the whole energy of the tore would be lost in travelling one scale-height if material passing through the tore is lifted or depressed, on an average, 300 km.

At first sight it may appear possible to overcome this energy-difficulty by supposing that material travels exactly once (or a whole number of times) round the tore, and is left by the tore exactly in its original position. A motion of the type postulated would, however, be unstable; if the distance travelled became a little less than a complete revolution, dissipation of energy would begin, and would steadily increase as the distance was yet further reduced. Moreover, the gravitational forces against which work has to be done are not small compared with the forces maintaining the m.-h. waves, and, though an m.-h. wave could return each element of mass to its starting-point if the gravitational forces were inoperative, it is idle to hope that the same can occur when they are taken into account. The conclusion appears inevitable that, if the motions in a tore are not horizontal, there is a considerable dissipation of m.-h. energy; the energy dissipated is presumably that of the vertical motion which gives rise to the dissipation. Hence a tore reaching the surface must be very nearly in a horizontal plane, and cannot, therefore, produce a pair of spots.

(vii) The motion illustrated in Fig. 3 is contrary to Law II of Paragraph 2, for at the surface it implies that material is moving across the lines of force. Walén in fact does not regard Fig. 3 (which is similar to his Fig. 12) as giving anything like an exact representation of the field and motion near the surface, but I cannot see that he avoids this difficulty. As is clear from the discussion in Paragraph 6, in any region where m.-h. waves produce a large magnetic field, they must also produce a large motion, nearly parallel to the field. A field can pass into outer space, as shown in Fig. 3, only if the m.-h. motions abruptly cease when the tore has reached the surface. It is difficult to see why this should be so.

Walén actually supposes that, as the tore approaches the surface, the stream-lines of flow begin to leave the upper part of the tore, crossing the lines of force, and closing below the tore (see his Figs. 14 and 19*b*). There is no justification whatever for this; motion must always accompany, and be accompanied by, a disturbance magnetic field. In any case, he bases this supposition on formulae (his 43–46) derived on the express assumption that the tore is horizontal and that the density does not vary greatly in a distance comparable with the tore's cross-sectional width.

8. To sum up our criticisms, Alfvén's theory is rather incomplete and not always self-consistent. Alfvén does not explain fully how the magnetic field causes a spot; but if it is by the motion which it produces, the base of the spot must be at too great a depth, and the pressure-difference between the spot and the surrounding material must be too great, for the theory to be satisfactory. While

* The total m.-h. energy is the sum of material kinetic energy, and of magnetic energy, equal in amount.

this difficulty appears to justify rejecting the theory, it would perhaps be well not to be too absolute, since all theories ascribing the origin of sunspots to the effects of motion are open to similar difficulties. However, Alfvén's theory meets worse difficulties than others in explaining the Evershed velocities.

Regarding the explanation of the sunspot magnetic field in terms of magneto-hydrodynamic waves, the position seems still more adverse to the theory. The theory is vague just where it needs to be explicit—in describing the behaviour of the wave near the surface; there are strong reasons why m.-h. waves should not produce dark spots, particularly spots with the observed properties of sunspots. While one cannot but admire the fertility of Alfvén's ideas, and Walén's methodical development of them into an orderly whole, the objections to the theory appear to me altogether insuperable. Alfven's waves explain how a magnetic field like the Sun's recovers from a distortion due to convection, and incidentally show clearly how difficult observations of the Sun's general magnetic field must be in the disturbed conditions of sunspot maximum, quite apart from the disturbance due to spot field. But they cannot provide a theory of sunspots.

I should like to express my gratitude to Dr V. C. A. Ferraro, who read an earlier version of this paper, and made several useful suggestions. I have also benefited by an interesting correspondence with Dr Walén, which, though it did not convince me of the soundness of his point of view, enabled me to correct several inaccuracies in my own.

University College of North Wales,
Bangor:
1946 *September* 24.

THE DECAY OF THE GRAVITATIONAL ENERGY OF A DOUBLE STAR

G. L. Clark

(Received 1946 September 9 ; Addendum received 1946 October 7)

Summary

For twenty years it has been known that there is a decay of energy of a rotating cohesive system but hitherto the corresponding problem of a double star has defied solution. In this paper, by using the method previously employed by the author in discussing a rotating rod, it is shown that a double star does lose gravitational energy and that the radiation is of the same order of magnitude as for cohesive systems.

---

1. *Introduction.*—In a paper published in 1941 (paragraph 11) * and in another paper in the press † the author has discussed the problem of the loss of energy of a rotating cohesive system. By calculating the external field of the system he has obtained the angular equation of motion as a condition of integrability of the equations $G_{\mu\nu}=0$. Previous investigators ‡ had either used the pseudo energy-tensor or had made assumptions regarding the energy-tensor and none of the discussions had been free from criticism.

On the other hand the analogous problem of a double star has hitherto been unsolved. Only approximate solutions of the two-body problem, giving the accelerations correct to order m^2 have been obtained by calculating the value of the field in the neighbourhood of one of the singularities. § The rate of loss of energy of a rotating cohesive system is of order $(m/b)^2v^6$, where b is of the order of magnitude of the linear dimensions of the system. For gravitational systems m is of the order bv^2 or $b^3\omega^2$ so that the rate of loss of energy (if any) may be expected to be of the order $(m/b)^5$. The problem is too complicated to solve by merely attempting to extend the solutions quoted above and an entirely new argument is clearly required. The method which I used for discussing cohesive systems opens up the most promising line of attack. We accordingly calculate the field at some distance from the system instead of its value in the neighbourhood of one of the bodies and use retarded instead of present values. It is found that the field is identical with that of a rotating cohesive system and consequently there is also a loss of energy of the same magnitude. For the present we only consider two particles of equal mass moving in circular orbits about their centre of mass and we assume the relation between m, b and ω which is obtained as a condition for empty space when we calculate the field in the neighbourhood of one of the particles.

* G. L. Clark, *Proc. Roy. Soc.*, A, **177,** 245-248, 1941.

† G. L. Clark, *Proc. Camb. phil. Soc.* (in press).

‡ A. S. Eddington, *Proc. Roy. Soc.*, A, **102,** 268-282, 1922 ; A. S. Eddington, *Mathematical theory of relativity*, p. 248, 2nd edition, Cambridge, 1924 ; A. Einstein, *S.B. preuss. Akad. Wiss.*, 688, 1916 ; A. Einstein, *S.B. preuss. Akad. Wiss.*, 154, 1918.

§ A. S. Eddington and G. L. Clark, *Proc. Roy. Soc.*, A, **166,** 465-475, 1938 ; A. Einstein, L. Infeld and B. Hoffmann, *Ann. Math. Princeton*, etc., **39,** 65-100, 1938 ; H. P. Robertson, *Ann. Math. Princeton*, etc., **39,** 101-104, 1938 ; G. L. Clark, *Proc. Roy. Soc.*, A, **177,** 240-245, 1941.

2. *The terms linear in the density.*—In the paper on rotating systems * I have found it convenient to use potentials $\gamma_{\mu\nu}$ instead of the $g_{\mu\nu}$, where

$$\gamma_{\mu\nu} = h_{\mu\nu} - \tfrac{1}{2}\eta_{\mu\nu}\eta^{\sigma\rho}h_{\sigma\rho}, \tag{2.1}$$

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{2.2}$$

and the $\eta_{\mu\nu}$ represent the Galilean values of the $g_{\mu\nu}$. By using the convention that Greek indices take the values 1, 2, 3 and 4 while Latin indices take the values 1, 2, 3 only, the linear terms in the field equations may be expressed in the form

$$\left.\begin{aligned} 16\pi T_{mn} &= \gamma_{mn|ss} - \gamma_{ns|sm} - \gamma_{mn|44} + \gamma_{4m|4n} + \gamma_{4n|4m} \\ &\quad + \delta_{mn}(\gamma_{rs|rs} - 2\gamma_{4s|4s} + \gamma_{44|44}), \\ 16\pi T_{4n} &= \gamma_{4n|ss} - \gamma_{ns|4s} + \gamma_{44|4n} + \gamma_{4s|sn}, \\ 16\pi T_{44} &= \gamma_{44|ss} - \gamma_{rs|sr}, \end{aligned}\right\} \tag{2.3}$$

where a line followed by a suffix denotes ordinary differentiation and the coordinate x_4 represents the time. If we impose coordinate conditions

$$\gamma_{4\mu|4} - \gamma_{\mu s|s} = 0, \tag{2.4}$$

the equations (2.3) take the form

$$16\pi T_{\mu\nu} = \gamma_{\mu\nu|ss} - \gamma_{\mu\nu|44}. \tag{2.5}$$

In the case of a simple source the solution of (2.5) given by Eddington † is

$$\gamma_{\mu\nu} = -4\{T_{\mu\nu}\}_{t-r}r^{-1}, \tag{2.6}$$

where $\{T_{\mu\nu}\} = \int T_{\mu\nu}dV$ taken over the source, $\{\}_{t-r}$ indicates that retarded values are used and $t = x_4$.

For a double star having components of equal mass m, the axes of reference may be chosen so that the coordinates of the components relative to the centre of mass at the Newtonian level are

$$(b\cos\omega t,\ b\sin\omega t,\ 0) \text{ and } (-b\cos\omega t,\ -b\sin\omega t,\ 0), \tag{2.7}$$

where b and ω are constants and the component velocities are

$$(-b\omega\sin\omega t,\ b\omega\cos\omega t,\ 0) \text{ and } (b\omega\sin\omega t,\ -b\omega\cos\omega t,\ 0). \tag{2.8}$$

The antedated values of the velocity for an observer at (x_1, x_2, x_3) are accordingly

$$(-b\omega\sin\theta_1,\ b\omega\cos\theta_1,\ 0) \text{ and } (b\omega\sin\theta_2\ -b\omega\cos\theta_2,\ 0), \tag{2.9}$$

where

$$\left.\begin{aligned} r_1^2 &= R^2 - 2b(x_1\cos\theta_1 + x_2\sin\theta_1) + b^2, \\ r_2^2 &= R^2 + 2b(x_1\cos\theta_2 + x_2\sin\theta_2) + b^2, \\ R^2 &= x_1^2 + x_2^2 + x_3^2, \\ \theta_1 &= \omega(t - r_1), \\ \theta_2 &= \omega(t - r_2). \end{aligned}\right\} \tag{2.10}$$

As far as the terms linear in the density are concerned we shall split up the potentials into two parts, thus

$$\gamma_{\mu\nu} = (\gamma_{\mu\nu})_1 + (\gamma_{\mu\nu})_2, \tag{2.11}$$

where the subscripts 1 and 2 respectively refer to the first and second singularity. We note at the outset that $(\gamma_{\mu\nu})_2$ can be deduced from $(\gamma_{\mu\nu})_1$ merely by changing

* G. L. Clark, *Proc. Camb. phil. Soc.* (in press).

† *Mathematical theory of relativity, loc. cit.*; *Proc. Roy. Soc.*, A, **102**, p. 274, 1922.

the sign of b. We begin therefore by ignoring the interaction between the particles and consider each separately. At very great distances from the system we see by (2.6) and (2.9) that the leading terms in $(\gamma_{41})_1$ and $(\gamma_{42})_1$ are, neglecting b^2,

$$(\gamma_{41})_1 = -4mb\omega \sin\theta_1 r_1^{-1}, \quad (\gamma_{42})_1 = 4mb\omega \cos\theta_1 r_1^{-1}. \tag{2.12}$$

Similarly the leading term in $(\gamma_{44})_1$ is

$$(\gamma_{44})_1 = -4mr^{-1}. \tag{2.13}$$

By adding terms containing higher powers of b, I have built up potentials satisfying

$$(\gamma_{\mu\nu|ss})_1 - (\gamma_{\mu\nu|44})_1 = 0 \tag{2.14}$$

and

$$(\gamma_{4\mu|4})_1 - (\gamma_{\mu s|s})_1 = \frac{5mb\omega^3}{1-B}(x_2\delta_{\mu 1} - x_1\delta_{\mu 2})r_1^{-4}, \tag{2.15}$$

where

$$B = b\omega(-x_1 \sin\theta_1 + x_2 \cos\theta_1)r_1^{-1} \tag{2.16}$$

and terms of order -4 in r_1 are neglected. On substituting (2.14) and (2.15) in (2.3) we find that $T_{\mu\nu}$ is of order -4 in r_1. That is, our solution satisfies the condition for empty space provided terms of order -4 in r_1 are neglected. Also, it has already been remarked that by calculating the field in the neighbourhood of one of the particles we obtain a relation between m, b and ω, which on the classical level gives

$$m \sim b^3\omega^2. \tag{2.17}$$

For the moment, however, we make no assumptions regarding the order of magnitude of b or ω. The potentials satisfying (2.14) and (2.15) which I have obtained are (on writing $p = 2\omega$ and dropping the subscript 1 to avoid unnecessary printing) *

$$\gamma_{44} = \gamma_{44}^{(1)} + \gamma_{44}^{(2)}, \quad \gamma_{4n} = \gamma_{4n}^{(1)} + \gamma_{4n}^{(2)}, \quad \gamma_{mn} = \sum_{r=1}^{5} \gamma_{mn}^{(r)}, \tag{2.18}$$

where

$$\begin{aligned}\gamma_{44}^{(1)} = &-4mr^{-1} + 4mb\omega(x_1 \sin\theta - x_2\cos\theta)r^{-2} - \tfrac{1}{2}mb^2p^2(x_1^2 + x_2^2)r^{-3} \\ &+ \tfrac{3}{2}mb^2p\{(x_2^2 - x_1^2)\sin 2\theta + 2x_1x_2\cos 2\theta\}r^{-4} \\ &+ \tfrac{3}{2}mb^2\{(x_1^2 - x_2^2)\cos 2\theta + 2x_1x_2\sin 2\theta\}r^{-5} - mb^3p^2(x_1\cos\theta + x_2\sin\theta)r^{-3} \\ &+ 3mb^3p^2(x_1^2 + x_2^2)(x_1\cos\theta + x_2\sin\theta)r^{-5},\end{aligned} \tag{2.19}$$

$$\left.\begin{aligned}\gamma_{41}^{(1)} = &-4mb\omega\sin\theta r^{-1} + 2mb^2\omega^2x_1r^{-2} + \tfrac{1}{2}mb^2p(x_1\sin 2\theta - x_2\cos 2\theta)r^{-3} \\ &- \tfrac{3}{2}mb^3p^2x_1(x_1\cos\theta + x_2\sin\theta)r^{-4}, \\ \gamma_{42}^{(1)} = &\,4mb\omega\cos\theta r^{-1} + 2mb^2\omega^2x_2r^{-2} - \tfrac{1}{2}mb^2p(x_1\cos 2\theta + x_2\sin 2\theta)r^{-3} \\ &- \tfrac{3}{2}mb^3p^2x_2(x_1\cos\theta + x_2\sin\theta^3r^{-4}, \\ \gamma_{43}^{(1)} = &\,0,\end{aligned}\right\} \tag{2.20}$$

$$\left.\begin{aligned}\gamma_{11}^{(1)} &= 4mb\omega(x_1\sin\theta + x_2\cos\theta)r^{-2} + 4\,mb(-x_1\cos\theta + x_2\sin\theta)r^{-3}, \\ \gamma_{12}^{(1)} &= 4mb\omega(-x_1\cos\theta + x_2\sin\theta)r^{-2} - 4mb(x_1\sin\theta + x_2\cos\theta)r^{-3}, \\ \gamma_{22}^{(1)} &= -\gamma_{11}^{(1)}, \\ \gamma_{13}^{(1)} &= 4mb\omega x_3\sin\theta r^{-2} - 4mbx_3\cos\theta r^{-3}, \\ \gamma_{23}^{(1)} &= -4mbx_3\cos\theta r^{-2} - 4mbx_3\sin\theta r^{-3}, \\ \gamma_{33}^{(1)} &= 4mb\omega(-x_1\sin\theta + x_2\cos\theta)r^{-2} + 4mb(x_1\cos\theta + x_2\sin\theta)r^{-3},\end{aligned}\right\} \tag{2.21}$$

* We use p or 2ω indifferently and assume ω to be constant. Also throughout the rest of the paper we denote r_1, θ_1 by r, θ unless otherwise stated.

$$\left.\begin{aligned}
\gamma^{(2)}_{11} &= 4mb^2\cos 2\theta r^{-3} - \tfrac{5}{2}mb^2p\sin 2\theta r^{-2} + \tfrac{1}{2}mb^2p^2(x_2^2 - x_1^2)r^{-3} - 6mb^2px_1x_2r^{-4},\\
\gamma^{(2)}_{12} &= 4mb^2\sin 2\theta r^{-3} + \tfrac{5}{2}mb^2p\cos 2\theta r^{-2} - mb^2p^2x_1x_2r^{-3} + 3mb^2p(x_1^2 - x_2^2)r^{-4},\\
\gamma^{(2)}_{22} &= -\gamma^{(2)}_{11},\\
\gamma^{(2)}_{13} &= -\tfrac{1}{2}mb^2p^2x_1x_3r^{-3} - 3mb^2px_2x_3r^{-4},\\
\gamma^{(2)}_{23} &= -\tfrac{1}{2}mb^2p^2x_2x_2r^{-3} + 3mb^2px_1x_3r^{-4},\\
\gamma^{(2)}_{33} &= \tfrac{1}{2}mb^2p^2(x_1^2 + x_2^2)r^{-3},
\end{aligned}\right\} \quad (2.22)$$

$$\left.\begin{aligned}
\gamma^{(3)}_{11} &= \tfrac{1}{2}mb^2p^2x_3^2\cos 2\theta r^{-3} + mb^2p\{\tfrac{5}{2}r^{-2} - 3(x_1^2 + x_2^2)r^{-4}\}\sin 2\theta\\
&\quad + mb^2\{6(x_1^2 + x_2^2)r^{-5} - 4r^{-3}\}\cos 2\theta,\\
\gamma^{(3)}_{12} &= \tfrac{1}{2}mb^2p^2x_3^2\sin 2\theta r^{-3} + mb^2p\{3(x_1^2 + x_2^2)r^{-4} - \tfrac{5}{2}r^{-2}\}\cos 2\theta\\
&\quad + mb^2\{6(x_1^2 + x_2^2)r^{-5} - 4r^{-3}\}\sin 2\theta,\\
\gamma^{(3)}_{22} &= -\gamma^{(3)}_{11},\\
\gamma^{(3)}_{13} &= -\tfrac{1}{2}mb^2p^2x_3(x_1\cos 2\theta + x_2\sin 2\theta)r^{-3} + 3mb^2px_3(-x_1\sin 2\theta\\
&\quad + x_2\cos 2\theta)r^{-4} + 6mb^2x_3(x_1\cos 2\theta + x_2\sin 2\theta)r^{-5},\\
\gamma^{(3)}_{23} &= \tfrac{1}{2}mb^2p^2x_3(-x_1\sin 2\theta + x_2\cos 2\theta)r^{-3} + 3mb^2px_3(x_1\cos 2\theta + x_2\sin 2\theta)r^{-4}\\
&\quad + 6mb^2x_3(x_1\sin 2\theta - x^2\cos 2\theta)r^{-5},\\
\gamma^{(3)}_{33} &= \tfrac{1}{2}mb^2p^2\{(x_1^2 - x_2^2)\cos 2\theta + 2x_1x_2\sin 2\theta\}r^{-3} + 3mb^2p\{(x_1^2 - x_2^2)\sin 2\theta\\
&\quad - 2x_1x_2\cos 2\theta\}r^{-4} + 6mb^2\{(x_2^2 - \tfrac{2}{1})\cos 2\theta - 2x_1x_2\sin 2\theta\}r^{-5},
\end{aligned}\right\} \quad (2.23)$$

$$\left.\begin{aligned}
\gamma^{(4)}_{11} &= \tfrac{3}{2}mb^3p^2x_3^2(x_1\cos\theta - x_2\sin\theta)r^{-5} + 3mb^3p^2(x_1^2 - x_2^2)(x_1\cos\theta + x_2\sin\theta)r^{-5}\\
&\quad + \tfrac{1}{4}mb^3p^2(-x_1\cos\theta + x_2\sin\theta)r^{-3},\\
\gamma^{(4)}_{22} &= -\gamma^{(4)}_{11},\\
\gamma^{(4)}_{12} &= \tfrac{3}{2}mb^3p^2(x_2\sin\theta + x_2\cos\theta)r^{-5} + 6mb^3p^2x_1x_2(x_1\cos\theta + x_2\sin\theta)r^{-5}\\
&\quad + \tfrac{1}{4}mb^3p^2(-x_1\sin\theta - x_2\cos\theta)r^{-3},\\
\gamma^{(4)}_{13} &= \tfrac{3}{2}mb^3p^2x_3\{(x_2^2 - x_1^2)\cos\theta + 2x_1x_2\sin\theta\}r^{-5} + mb^3p^2x_3\cos\theta r^{-3},\\
\gamma^{(4)}_{23} &= \tfrac{3}{2}mb^3p^2x_3\{(x_2^2 - x_1^2)\sin\theta + 2x_1x_2\cos\theta\}r^{-5} + mb^3p^2x_3\sin\theta r^{-3},\\
\gamma^{(4)}_{33} &= \tfrac{3}{2}mb^3p^2x_3^2(x_1\cos\theta + x_2\sin\theta)r^{-5} - \tfrac{5}{2}mb^3p^2(x_1\cos\theta + x_2\sin\theta)r^{-3},
\end{aligned}\right\} \quad (2.24)$$

$$\left.\begin{aligned}
\gamma^{(5)}_{11} &= -\tfrac{1}{2}mb^3p^2x_3^2(x_1\cos 3\theta + x_2\sin 3\theta)r^{-5} + \tfrac{1}{12}mb^3p^2(x_1\cos 3\theta\\
&\quad + x_2\sin 3\theta)r^{-3},\\
\gamma^{(5)}_{22} &= -\gamma^{(5)}_{11},\\
\gamma^{(5)}_{12} &= \tfrac{1}{2}mb^3p^2x_3^2(-x_1\sin 3\theta + x_2\cos 3\theta)r^{-5} + \tfrac{1}{12}mb^3p^2(x_1\sin 3\theta - x_2\cos 3\theta)r^{-3},\\
\gamma^{(5)}_{13} &= \tfrac{1}{2}mb^3p^2x_3\{(x_1^2 - x_2^2\cos 3\theta + 2x_1x_2\sin 3\theta\}r^{-5},\\
\gamma^{(5)}_{23} &= \tfrac{1}{2}mb^3p^2x_3\{(x_1^2 - x_2^2)\sin 3\theta - 2x_1x_2\cos 3\theta\}r^{-5},\\
\gamma^{(5)}_{33} &= \tfrac{1}{2}mb^3p^2\{x_1(3x_2^2 - x_1^2)\cos 3\theta + x_2(x_2^2 - 3x_1^2)\sin 3\theta\}r^{-5},
\end{aligned}\right\} \quad (2.25)$$

and

$$\left.\begin{aligned}
\gamma^{(2)}_{44} &= \sum_{j=1}^{3}\frac{mb^3\omega^2}{j}\left\{\left(-\phi^{(j)}_4 + \phi^{(j)}_s\frac{x_s}{r}\right)\sin j\theta + \left(\psi^{(j)}_4 - \psi^{(j)}_s\frac{x_s}{r}\right)\cos j\theta\right\},\\
\gamma^{(2)}_{45} &= \sum_{j=1}^{3}\frac{mb^3\omega^2}{j}(-\phi^{(j)}_s\sin j\theta + \psi^{(j)}_s\cos j\theta),
\end{aligned}\right\} \quad (2.26)$$

where $\phi^{(j)}_\mu$ and $\psi^{(j)}_\mu$ are of order -3 in r and the factor $b^3\omega^2$ indicates that the terms of order 3 in b and 2 in ω at least. The actual values of $\phi^{(j)}_\mu$ and $\psi^{(j)}_\mu$ are not required in this investigation.

3. *The field in terms of R at great distances.*—In the preceding section we have written down potentials containing two singularities for which the linear terms in the expressions for $T_{\mu\nu}$ identically vanish provided terms of order -4 in

and r_2 are neglected. The next step in the analysis is to substitute these potentials in the non-linear terms of $G_{\mu\nu}$. For the sake of simplicity we shall make use of the condition (2.17) and neglect contributions from the terms of order 3 in b. Also we shall no longer distinguish between the singularities but express r_1 and r_2 in terms of R and b.

To the second order in b we write

$$r_1 = R(1 + bs_1 + b^2 s_2). \tag{3.1}$$

We then have correct to the first order in b,

$$\left.\begin{aligned} \cos\omega(t-r_1) &= \cos\omega(t-R) + \omega R b s_1 \sin\omega(t-R), \\ \sin\omega(t-r_1) &= \sin\omega(t-R) - \omega R b s_1 \cos\omega(t-R). \end{aligned}\right\} \tag{3.2}$$

Substituting (3.1) and (3.2) in (2.10) and equating the coefficients of b and b^2 gives respectively

$$s_1 = -x_1 R^{-2}\cos\omega(t-R) - x_2 R^{-2}\sin\omega(t-R) \tag{3.3}$$

and

$$s_2 = \tfrac{1}{2}R^{-2} - \tfrac{1}{2}s_1^2 + \omega s_1 R^{-1}\{-x_1\sin\omega(t-R) + x_2\cos\omega(t-R)\}. \tag{3.4}$$

We now substitute (3.2), (3.3) and (3.4) in the potentials (2.19)–(2.23) and add the corresponding expressions for $(\gamma_{\mu\nu})_2$. The result, on retaining terms of order -3 in R in the case of periodic terms and -2 in R in the case of non-periodic terms, is

$$\left.\begin{aligned}
\gamma_{44} &= -8mR^{-1} + Ip^2R^{-3}\{(x_1^2 - x_2^2)\cos p(t-R) + 2x_1x_2\sin p(t-R)\} \\
&\quad + 3IpR^{-4}\{(x_1^2 - x_2^2)\sin p(t-R) - 2x_1x_2\cos p(t-R)\} \\
&\quad - 3IR^{-5}\{(x_1^2 - x_2^2)\cos p(t-R) + 2x_1x_2\sin p(t-R)\}, \\
\gamma_{41} &= -2A\omega x_2R^{-3} - Ip^2R^{-2}\{x_1\cos p(t-R) + x_2\sin p(t-R)\} \\
&\quad + IpR^{-3}\{-x_1\sin p(t-R) + x_2\cos p(t-R)\}, \\
\gamma_{42} &= 2A\omega x_1R^{-3} + Ip^2R^{-2}\{-x_1\sin p(t-R) + x_2\cos p(t-R)\} \\
&\quad + IpR^{-3}\{x_1\cos p(t-R) + x_2\sin p(t-R)\}, \\
\gamma_{11} &= -\gamma_{22} = Ip^2R^{-1}\cos p(t-R), \\
\gamma_{12} &= Ip^2R^{-1}\sin p(t-R),
\end{aligned}\right\} \tag{3.5}$$

and the remaining components of $\gamma_{\mu\nu}$ are zero to the order considered. In (3.5) $I = mb^2$ and $A = 2I$.

The angular equation of motion is obtained by considering the non-linear terms $2L_{\mu\nu}$ in the expression for $G_{\mu\nu}$; the particular terms which we consider being the non-periodic terms of order -3 in R. Now the potentials (3.5) are of the form

$$\phi = mf_1 + mIp^2 f_2 \begin{matrix}\sin \\ \cos\end{matrix} p(t-R), \tag{3.6}$$

where f_1, f_2 are of the order -1 in R and typical partial contributions to $2L_{\mu\nu}$ from the terms quadratic in $\gamma_{\mu\nu}$ are

$$\phi\phi_{|\mu\nu} \quad \text{and} \quad \phi_{|\mu}\phi_{|\nu}. \tag{3.7}$$

It is clear therefore that the non-periodic terms of order -3 in R from the terms quadratic in $\gamma_{\mu\nu}$ are of order I^2p^5. On the other hand typical partial contributions to $2L_{\mu\nu}$ from the terms cubic in $\gamma_{\mu\nu}$ are

$$\phi\phi\phi_{|\mu\nu} \quad \text{and} \quad \phi\phi_{|\mu}\phi_{|\nu}. \tag{3.8}$$

The non-periodic terms of order -3 in R in $2L_{\mu\nu}$ from the cubic terms are

therefore of order

$$mI^2p^6 \sim m^3b^4\omega^6 \sim m^2b^7\omega^8 \sim mb^{10}\omega^{10}, \tag{3.9}$$

on using the classical result (2.17). The contribution from the terms cubic in $\gamma_{\mu\nu}$ may therefore be neglected and we confine our attention to the terms arising from the quadratic terms only.

The result (3.5) is extremely important as the potentials are of precisely the same form as those at great distances from a rotating rod, except that in the case of a rod $A = I$ and not $2I$. In my (1946) discussion of a rotating cohesive system I show, by modifying my 1941 investigation that we have to abandon the assumption that ω is constant and that the equation of motion is *

$$A\dot{\omega} + \frac{32}{5} I^2\omega^5 = 0. \tag{3.10}$$

Consequently in the case of a double star as well as in the case of a rod, there is a decay of gravitational energy of order $m^2b^4\omega^6$ or m^2v^6/b^2.

In conclusion we remark that when powers of b greater than 2 are retained in the terms linear in the density, the expressions for $(\gamma_{\mu\nu})_1$ and $(\gamma_{\mu\nu})_2$ in terms of R are respectively of the form

$$\left.\begin{aligned}(\gamma_{\mu\nu})_1 &= mb\chi_1 + mb^2\chi_2 + mb^3\chi_3 + mb^4\chi_4 + \ldots, \\ (\gamma_{\mu\nu})_2 &= -mb\chi_1 + mb^2\chi_2 - mb^3\chi_3 + mb^4\chi_4 - \ldots.\end{aligned}\right\} \tag{3.11}$$

Hence

$$\gamma_{\mu\nu} = (\gamma_{\mu\nu})_1 + (\gamma_{\mu\nu})_2 = 2mb^2\chi_2 + 2mb^4\chi_4 + 2mb^6\chi_6 + \ldots, \tag{3.12}$$

and the $\gamma_{\mu\nu}$ will accordingly consist of even powers of b only. In particular no terms of order mb^7 occur, and therefore no terms of order $I^2\omega^5$ can possibly come into the equation of motion from the terms linear in the density. The whole purpose of giving the terms of order 3 in b in the second section of this paper is to verify that no terms of order $mb^4\omega^4$ or $mb^6\omega^6$ appear in the equation of motion.

Addendum.—In Section 2 we built up expressions for the external field due to one of the moving singularities but we did not take precautions to ensure that the field was due to a *spherical* particle whose density was a function of the distance from its centre.† We adopted the method used in order to give a theoretically simple discussion which was at the same time perfectly rigorous. The result obtained, is however, generally true and, in particular, in the case of spherical particles whose density is symmetrical about their respective centres.

The solution for γ_{4n} may be written down at the outset; namely,

$$\gamma_{4n} = -[4T_{4n}/r(1 - v_r)], \tag{1}$$

where the bracket denotes that the quantities are antedated and v_r is the component of the velocity in the direction of r.‡ Now in the case of the first particle,

$$\begin{aligned} v_r &= b\omega\{-(x_1 - b\cos\theta_1)\sin\theta_1 + (x_2 - b\sin\theta_1)\cos\theta_1\}r_1^{-1} \\ &= b\omega(-x_1\sin\theta_1 + x_2\cos\theta_1)r_1^{-1}. \end{aligned} \tag{2}$$

---

* In all cases A is the moment of inertia.

† I suspect that the field (2.22)-(2.26) would have to be considered as instantaneous (i.e. t must be equated to zero after all differentiations have been carried out) for the solution to represent the field of a physically interesting distribution of matter inside the sphere.

‡ Eddington, *Mathematical theory of relativity*, p. 249 equation (6), Cambridge, 1924.

We also take

$$T_{41} = 4mb\omega \sin\theta_1, \quad T_{42} = -4mb\omega\cos\theta_1, \quad T_{43} = 0. \tag{3}$$

Substituting (3) in (1) and neglecting cubes and higher powers of b gives,

$$(\gamma_{41})_1 = -4mb\omega\sin\theta_1 r_1^{-1} + 2mb^2\omega^2 x_1 r_1^{-2} - 2mb^2\omega^2(x_1\cos 2\theta_1 + x_2\sin 2\theta_1)r_1^{-2}. \tag{4}$$

Also we have

$$r_1 = R(1 + bs_1), \tag{5}$$

where

$$s_1 = -(x_2\cos\theta_1 + x_2\sin\theta_1)R^{-2}. \tag{6}$$

On expanding $(\gamma_{41})_1$ in powers of $1/R$, we find on neglecting b^3,

$$\begin{aligned}(\gamma_{41})_1 = {} & -4mb\omega\sin\omega(t-R)R^{-1} - 2mb^2\omega x_2 R^{-3} + 2mb^2\omega R^{-3}\{-x_1\sin p(t-R) \\ & + x_2\cos p(t-R)\} - 4mb^2\omega^2 R^{-2}\{x_1\cos p(t-R) + x_2\sin p(t-R)\}.\end{aligned} \tag{7}$$

Hence,

$$\begin{aligned}\gamma_{41} &= (\gamma_{41})_1 + (\gamma_{41})_2 \\ &= -2A\omega x_2 R^{-3} - Ip^2R^{-2}\{x_1\cos p(t-R) + x_2\sin p(t-R)\} \\ &\qquad + IpR^{-3}\{-x_1\sin p(t-R) + x_2\cos p(t-R)\}.\end{aligned} \tag{8}$$

Similarly, we have

$$\begin{aligned}\gamma_{42} = {} & 2A\omega x_1 R^{-3} + Ip^2R^{-2}\{-x_1\sin p(t-R) + x_2\cos p(t-R)\} \\ & + IpR^{-3}\{x_1\cos p(t-R) + x_2\sin p(t-R)\}.\end{aligned} \tag{9}$$

These expressions for γ_{4n} are precisely those given in (3.5). In a similar manner the value of γ_{44} is obtained by expanding

$$\begin{aligned}(\gamma_{44})_1 &= -[4m/r(1-v_r)]_1 \\ &= -4mr_1^{-1} + 4mb\omega r_1^{-2}(x_1\sin\theta_1 - x_2\cos\theta_1) - 2mb^2\omega^2(x_1^2 + x_2^2)r_1^{-3} \\ &\quad + 2mb^2\omega^2 r_1^{-3}\{(x_1^2 - \tfrac{2}{2}x)\cos 2\theta_1 + 2x_1x_2\sin 2\theta_1\}.\end{aligned} \tag{10}$$

The components γ_{mn} must then necessarily take the values previously given.

Trinity College,
Cambridge :
1946 *September* 8 and *October* 6.

"SUDDEN COMMENCEMENTS" IN THE GREENWICH MAGNETIC RECORDS (1879-1944) AND RELATED SUNSPOT DATA*

H. W. Newton

(Communicated by the Astronomer Royal)
(Received 1946 August 27)

Summary

Part I: Statistics of "sudden commencements"—their average amplitude, direction of impulse, hourly and monthly frequencies—are derived from 681 cases identified on the Greenwich magnetograms over six sunspot cycles. Besides SCs that mark the onset of many geomagnetic storms, there are frequent cases in which no appreciable disturbance follows. About half of all SCs have a small preliminary impulse (lasting about half a minute) in opposition to the main impulse, most commonly of 3 minutes' duration. The presence of this preliminary impulse seems subject to some diurnal effect.

Both diurnal and seasonal effects appear in SC frequency occurrence. An unexpected result is a marked minimum around 8^h–9^h. SCs that do occur then are usually inverted (the normal main stroke is $+\Delta H$, $+\Delta V$ and $-\Delta D$). Details are tabulated for every SC recorded at Abinger from 1938 November to 1944 December for comparison elsewhere.

Part II: SC occurrence in relation to the sunspot cycle and to individual sunspots is investigated. Each of the four groups into which the data were divided gives a frequency curve similar to the mean sunspot curve. In this respect, small geomagnetic storms with no SC onset have a different distribution within the sunspot cycle. A marked tendency for SCs followed by storms to occur when big spots are about one day past the central meridian accords with a well-known relationship between the larger magnetic storms and the bigger sunspots. The association of SCs not followed by disturbance with respect to a big sunspot appears largely fortuitous, but some effect from spots nearing the western limb should not be ruled out.

An addendum gives the results of a sample comparison of SC pulses and other minor movements on the Abinger and Lerwick magnetograms. SCs as recorded at the two stations are very similar, but ΔV at Lerwick is normally negative. During inactive periods a remarkable similarity of even small detail is noted between these two stations which are over 600 miles apart.

* The full text of this paper will be published in *M.N., Geophys. Suppl.*, **5**, No. 6, 1946.

Note: If a paper appearing in the *Geophysical Supplement* is of direct astronomical, as well as geophysical, interest, the summary will in future be reprinted in the Monthly Notices.

GEOMAGNETIC "CROCHET" OCCURRENCE AT ABINGER 1936-1946 AND ALLIED SOLAR AND RADIO DATA*

H. W. Newton

(Received 1947 April 2)

Summary

The rather infrequent occurrence at Abinger of geomagnetic "crochets" (ultra-violet solar radiation effect) has been studied over 11 years, in conjunction with data of solar flares and radio fade-outs. The trace movements in H, D and V of 23 examples are reproduced, in addition to tabulated data for these and for 32 other cases of less certain identification. The general synchronization of flare, fade-out and crochet (when present) is noteworthy. Comparison is made of the amplitude ΔH and direction ΔD of the crochets with the normal diurnal inequality of H and D. A striking feature is the change in direction from East to West in D at about $10^{h}\cdot7$ apparent solar time, in conformity with the diurnal inequality in D. The correlation with the diurnal inequality in H is less obvious, but ΔH is of the correct sign (negative) over the range of daylight hours in which crochets occurred. There are indications of other diurnal and seasonal factors (other than a simple dependence upon the Sun's zenith distance) in the occurrence of crochets at Abinger. Some crochets in 1946 have synchronized with reported bursts of "solar noise".

* The full text of this paper will be published in *M.N., Geophys. Suppl.*, **5**, No. 6, 1946.